PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C07K 14/00 — A2

(11) International Publication Number: **WO 00/21985**

(43) International Publication Date: 20 April 2000 (20.04.00)

(21) International Application Number: PCT/IB99/01729

(22) International Filing Date: 13 October 1999 (13.10.99)

(30) Priority Data:
60/104,299 14 October 1998 (14.10.98) US

(71) Applicant *(for all designated States except US)*: GENSET [FR/FR]; 24, rue Royale, F–75008 Paris (FR).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: BOUGUELERET, Lydie [FR/FR]; 108, avenue Victor Hugo, F–92170 Vanves (FR). MALEKZADEH, Kattayoun [IR/FR]; 41, rue des Laitieres, F–94300 Vincennes (FR).

(74) Common Representative: GENSET; 24, rue Royale, F–75008 Paris (FR).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: GENES ENCODING OLFACTORY RECEPTORS AND BIALLELIC MARKERS THEREOF

(57) Abstract

The invention concerns the genomic sequence and coding regions of a new olfactory receptor gene cluster. The invention also concerns polypeptides encoded by the olfactory receptor genes as well as to methods and kits for detecting these polynucleotides and screening substances interacting with these polypeptides. The invention also deals with antibodies directed specifically against such polypeptides that are useful as diagnostic reagents. The invention further encompasses biallelic markers of the olfactory receptor gene useful in genetic analysis.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

GENES ENCODING OLFACTORY RECEPTORS AND BIALLELIC MARKERS THEREOF

FIELD OF THE INVENTION

The present invention pertains to a purified or isolated nucleic acid comprising ten open reading Frames (ORFs) encoding ten different olfactory receptor-like proteins, non-coding regions flanking the ORFs as well as fragments thereof. The invention also provides recombinant expression vectors and recombinant cell hosts containing a nucleic acid encoding said olfactory receptor proteins. The invention also concerns the olfactory receptor proteins encoded by these ORFs as well as polypeptides that are homologous to said olfactory receptor proteins and the peptide fragments of both the olfactory receptor proteins and their homologous polypeptide counterparts. The invention also deals with antibodies directed specifically against such polypeptides that are useful as diagnostic reagents. The invention further encompasses biallelic markers of the olfactory receptor gene useful in genetic analysis. The invention also deals with methods and kits for the detection of the olfactory receptor proteins and with methods and kits for screening ligand molecules binding to these proteins.

BACKGROUND OF THE INVENTION

Throughout this application, various bibliographic publications are cited. Full bibliographic references for these publications may be found at the end of this application, preceding the sequence listing and the claims.

OLFACTORY SYSTEM

The olfactory receptor cells, the first cells in the pathway that give rise to the sense of smell, lie in a small patch of membrane, the olfactory epithelium, in the upper part of the nasal cavity. These cells are specialized afferent neurons that have an enlarged extension analogous to a dendrite. Several long hairlike processes extend out from this extension along the surface of the olfactory epithelium where they are bathed in mucus. The hairlike processes contain the receptor proteins for olfactory stimuli. The axons of these neurons form the olfactory nerve.

For the detection of an odorous substance which is called an odorant, molecules of the substance must first diffuse into the air and pass into the nose to the region of the olfactory epithelium. Once there, they dissolve in the mucus that covers the epithelium and then bind to specific receptor proteins on the cilia.

Although there are many thousands of olfactory neurons, each contains one, or at most a few, of the 1,000 or so different receptor types, each of which responds only to a specific chemically related group of odorant molecules. Each odorant has characteristic chemical groups that distinguish it from other odorants, and each of these groups activates a different receptor type. Thus the identity

of a particular odorant is determined by the activation of a precise combination of receptors, each of which is contained in a distinct group of olfactory neurons.

The axons of the olfactory neurons synapse in the brain structures known as olfactory bulbs, which lie on the undersurface of the frontal lobes. Axons from olfactory neurons sharing a common receptor specificity synapse together on certain olfactory-bulb neurons, thereby maintaining the specificity of the original stimuli.

OLFACTORY RECEPTORS

In contrast with the immunoglobulin system, the diversity of olfactory receptors is encoded by a large germ-line repertoire of olfactory receptor genes. The size of the olfactory receptor gene family in the human genome is unknown but it has been estimated to encompass 200 to 1,000 genes.

The locations of only a few human genes have been determined to date. The picture that has emerged so far is that several large clusters of olfactory genes and pseudogenes span hundreds of kilobases on several chromosomes. Using FISH analyses, more than 25 distinct locations of olfactory receptors gene have been identified in the human genome.

In mammals, the olfactory epithelium appears to be organized into distinct topographic regions or zones in which expression of a particular receptor gene appears to be restricted to one of the four zones in the epithelium. Within the zone, the distribution of neurons expressing a given receptor is random. Chromosomal mapping studies have revealed clusters of odorant receptor genes at a single locus, and numerous such loci have been mapped to different chromosomes. However, receptors expressed in the same zone map to different loci, and a single locus can contain genes expressed in different zones. A putative odorant receptor promoter, consisting of the 6.7 kb DNA fragment upstream of the receptor coding region, has been shown to be sufficient to direct olfactory receptor expression in a tissue-specific, zonal-specific manner.

Olfactory receptors share a seven-transmembrane domain structure (TM1 to TM7) with many neurotransmitter and hormone receptors. They show a high degree of sequence similarity in some conserved domains (TM2 and TM7) as well as regions of diversity (TM3, TM4, TM5, and TM6). They are responsible for the recognition and G protein-mediated transduction of odorant signals. The genes encoding these receptors are devoid of introns within their coding regions.

Olfactory receptors display all hallmarks of the G-protein coupled receptor superfamily but have also some unique motifs. Most notably they appear to be minimal in structure with very short cytoplasmic and extracellular loops. In addition, they display a striking structural diversity in the third, fourth and fifth transmembrane domains which are supposed to form the hydrophobic core of these proteins, and may form the ligand binding site of the receptors.

An understanding of the genetic basis of olfaction and a knowledge of olfactory receptors are important to enable the design of fragrance, the identification of compounds which control appetite, or the detection of compounds which can be harmful or dangerous.

SUMMARY OF THE INVENTION

This invention provides a nucleic acid molecule encoding ten different olfactory receptor-like proteins (OLF).

The invention also deals with a nucleic acid molecule comprising a nucleotide sequence encoding an olfactory receptor-like protein, which nucleotide sequence is selected from the group consisting of SEQ ID Nos 2-11, as well as with the corresponding polypeptide encoded by this nucleotide sequence and with antibodies directed against the corresponding polypeptide.

Oligonucleotide probes or primers hybridizing specifically with an olfactory receptor genomic sequence are also part of the present invention, as well as DNA amplification and detection methods using said primers and probes.

The invention also concerns a purified and/or isolated biallelic marker located in the sequence of the olfactory receptor gene cluster of the invention, wherein said biallelic marker is useful as a diagnostic tool in order to detect an allele associated with a specific phenotype as regards to the olfaction system, including an alteration of the olfactory perception of substances or molecules.

A further object of the invention consists of recombinant vectors comprising any of the nucleic acid sequences described above, and in particular of recombinant vectors comprising a sequence encoding an olfactory receptor protein, as well as of cell hosts and transgenic non human animals comprising said nucleic acid sequences or recombinant vectors.

A further object of the invention consists of methods for screening substances or molecules interacting with an olfactory receptor encoded by any of the nucleic acid molecule described above.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 : Alignment of the amino acid sequences of the olfactory polypeptides encoded by the Open Reading Frames of the olfactory receptor gene cluster of the invention. The lower line represents the consensus sequence. The locations of the seven transmembrane domains TM1 to TM7 are boxed.

BRIEF DESCRIPTION OF THE SEQUENCES PROVIDED IN THE SEQUENCE LISTING

SEQ ID No 1 contains the olfactory receptor genomic sequence.

SEQ ID Nos 2-11 contains the nucleotide sequences of the open reading frame sequences of SEQ ID No 1 encoding the OLF1 to OLF10 polypeptides.

SEQ ID No 12-21 contain the amino acid sequence of OLF1 to OLF10 polypeptides encoded by the open reading frames of SEQ ID Nos 2-11.

SEQ ID Nos 22-25 contain the amplification primers used for FISH experiments described in Example 1.

SEQ ID No 26 contains a primer containing the additional PU 5' sequence described further in Example 3.

SEQ ID No 27 contains a primer containing the additional RP 5' sequence described further in Example 3.

In accordance with the regulations relating to Sequence Listings, the following codes have been used in the Sequence Listing to indicate the locations of biallelic markers within the sequences and to identify each of the alleles present at the polymorphic base. The code "r" in the sequences indicates that one allele of the polymorphic base is a guanine, while the other allele is an adenine. The code "y" in the sequences indicates that one allele of the polymorphic base is a thymine, while the other allele is a cytosine. The code "m" in the sequences indicates that one allele of the polymorphic base is an adenine, while the other allele is an cytosine. The code "k" in the sequences indicates that one allele of the polymorphic base is a guanine, while the other allele is a thymine. The code "s" in the sequences indicates that one allele of the polymorphic base is a guanine, while the other allele is a cytosine. The code "w" in the sequences indicates that one allele of the polymorphic base is an adenine, while the other allele is an thymine.

The nucleotide code of the original allele for each biallelic marker is the following:

Biallelic marker	Original allele
99-13670-305	G
99-13669-471	G
99-13666-275	A
99-13664-221	T
99-13663-218	G
99-13660-277	C
99-13652-407	G
99-13652-357	A
99-13652-308	A
99-13671-396	A
99-13649-286	C
99-13648-259	G
99-13647-278	G

DETAILED DESCRIPTION OF THE INVENTION

The aim of the present invention is to provide polynucleotides and polypeptides related to novel olfactory receptors, notably useful in order to design suitable means for detecting specific odorant molecules in a material sample, particularly in a material sample suspected to contain an odorant molecule that consists of one of the specific ligands for the olfactory receptors of the invention.

DEFINITIONS

Before describing the invention in greater detail, the following definitions are set forth to illustrate and define the meaning and scope of the terms used to describe the invention herein.

General definitions

The terms "olfactory receptor gene" or "OLF1 to OLF10" genes, when used herein, encompasses genomic, mRNA and cDNA sequences encoding the OLF1 to OLF10 olfactory receptor proteins.

The term "heterologous protein", when used herein, is intended to designate any protein or polypeptide other than the OLF1 to OLF10 proteins.

The term "isolated" requires that the material be removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or polypeptide present in a living animal is not isolated, but the same polynucleotide or DNA or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotide could be part of a vector and/or such polynucleotide or polypeptide could be part of a composition, and still be isolated in that the vector or composition is not part of its natural environment.

The term "purified" does not require absolute purity; rather, it is intended as a relative definition. Purification of starting material or natural material to at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated. As an example, purification from 0.1 % concentration to 10 % concentration is two orders of magnitude. The term "purified polynucleotide" is used herein to describe a polynucleotide or polynucleotide vector of the invention which has been separated from other compounds including, but not limited to other nucleic acids, carbohydrates, lipids and proteins (such as the enzymes used in the synthesis of the polynucleotide), or the separation of covalently closed polynucleotides from linear polynucleotides. A polynucleotide is substantially pure when at least about 50%, preferably 60 to 75% of a sample exhibits a single polynucleotide sequence and conformation (linear versus covalently close). A substantially pure polynucleotide typically comprises about 50%, preferably 60 to 90% weight/weight of a nucleic acid sample, more usually about 95%, and preferably is over about 99% pure. Polynucleotide purity or homogeneity is indicated by a number of means well known in the art, such as agarose or polyacrylamide gel electrophoresis of a sample, followed by visualizing a single polynucleotide band upon staining the gel. For certain purposes higher resolution can be provided by using HPLC or other means well known in the art.

The term "polypeptide" refers to a polymer of amino acids without regard to the length of the polymer; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide. This term also does not specify or exclude post-expression modifications of polypeptides, for example, polypeptides which include the covalent attachment of glycosyl groups, acetyl groups, phosphate groups, lipid groups and the like are expressly encompassed by the term

polypeptide. Also included within the definition are polypeptides which contain one or more analogs of an amino acid (including, for example, non-naturally occurring amino acids, amino acids which only occur naturally in an unrelated biological system, modified amino acids from mammalian systems etc.), polypeptides with substituted linkages, as well as other modifications known in the art, both naturally occurring and non-naturally occurring.

The term "recombinant polypeptide" is used herein to refer to polypeptides that have been artificially designed and which comprise at least two polypeptide sequences that are not found as contiguous polypeptide sequences in their initial natural environment, or to refer to polypeptides which have been expressed from a recombinant polynucleotide.

The term "purified polypeptide" is used herein to describe a polypeptide of the invention which has been separated from other compounds including, but not limited to nucleic acids, lipids, carbohydrates and other proteins. A polypeptide is substantially pure when at least about 50%, preferably 60 to 75% of a sample exhibits a single polypeptide sequence. A substantially pure polypeptide typically comprises about 50%, preferably 60 to 90% weight/weight of a protein sample, more usually about 95%, and preferably is over about 99% pure. Polypeptide purity or homogeneity is indicated by a number of means well known in the art, such as polyacrylamide gel electrophoresis of a sample, followed by visualizing a single polypeptide band upon staining the gel. For certain purposes higher resolution can be provided by using HPLC or other means well known in the art.

As used herein, the term "non-human animal" refers to any non-human vertebrate, birds and more usually mammals, preferably primates, farm animals such as swine, goats, sheep, donkeys, and horses, rabbits or rodents, more preferably rats or mice. As used herein, the term "animal" is used to refer to any vertebrate, preferable a mammal. Both the terms "animal" and "mammal" expressly embrace human subjects unless preceded with the term "non-human".

As used herein, the term "antibody" refers to a polypeptide or group of polypeptides which are comprised of at least one binding domain, where an antibody binding domain is formed from the folding of variable domains of an antibody molecule to form three-dimensional binding spaces with an internal surface shape and charge distribution complementary to the features of an antigenic determinant of an antigen, which allows an immunological reaction with the antigen. Antibodies include recombinant proteins comprising the binding domains, as wells as fragments, including Fab, Fab', $F(ab)_2$, and $F(ab')_2$ fragments.

As used herein, an "antigenic determinant" is the portion of an antigen molecule, in this case a OLF1 to OLF10 polypeptide, that determines the specificity of the antigen-antibody reaction. An "epitope" refers to an antigenic determinant of a polypeptide. An epitope can comprise as few as 3 amino acids in a spatial conformation which is unique to the epitope. Generally an epitope comprises at least 6 such amino acids, and more usually at least 8-10 such amino acids. Methods for determining the amino acids which make up an epitope include x-ray crystallography, 2-dimensional

nuclear magnetic resonance, and epitope mapping e.g. the Pepscan method described by Geysen et al. 1984; PCT Publication No. WO 84/03564; and PCT Publication No. WO 84/03506.

Throughout the present specification, the expression "nucleotide sequence" may be employed to designate indifferently a polynucleotide or a nucleic acid. More precisely, the expression "nucleotide sequence" encompasses the nucleic material itself and is thus not restricted to the sequence information (i.e. the succession of letters chosen among the four base letters) that biochemically characterizes a specific DNA or RNA molecule.

As used interchangeably herein, the terms "nucleic acids", "oligonucleotides", and "polynucleotides" include RNA, DNA, or RNA/DNA hybrid sequences of more than one nucleotide in either single chain or duplex form. The term "nucleotide" as used herein as an adjective to describe molecules comprising RNA, DNA, or RNA/DNA hybrid sequences of any length in single-stranded or duplex form. The term "nucleotide" is also used herein as a noun to refer to individual nucleotides or varieties of nucleotides, meaning a molecule, or individual unit in a larger nucleic acid molecule, comprising a purine or pyrimidine, a ribose or deoxyribose sugar moiety, and a phosphate group, or phosphodiester linkage in the case of nucleotides within an oligonucleotide or polynucleotide. The term "nucleotide" is also used herein to encompass "modified nucleotides" which comprise at least one modifications (a) an alternative linking group, (b) an analogous form of purine, (c) an analogous form of pyrimidine, or (d) an analogous sugar, for examples of analogous linking groups, purine, pyrimidines, and sugars see for example PCT publication No. WO 95/04064. The polynucleotide sequences of the invention may be prepared by any known method, including synthetic, recombinant, *ex vivo* generation, or a combination thereof, as well as utilizing any purification methods known in the art.

A "promoter" refers to a DNA sequence recognized by the synthetic machinery of the cell required to initiate the specific transcription of a gene.

A sequence which is "operably linked" to a regulatory sequence such as a promoter means that said regulatory element is in the correct location and orientation in relation to the nucleic acid to control RNA polymerase initiation and expression of the nucleic acid of interest. As used herein, the term "operably linked" refers to a linkage of polynucleotide elements in a functional relationship. For instance, a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the coding sequence. More precisely, two DNA molecules (such as a polynucleotide containing a promoter region and a polynucleotide encoding a desired polypeptide or polynucleotide) are said to be "operably linked" if the nature of the linkage between the two polynucleotides does not (1) result in the introduction of a frame-shift mutation or (2) interfere with the ability of the polynucleotide containing the promoter to direct the transcription of the coding polynucleotide.

The term "vector" is used herein to designate either a circular or a linear DNA or RNA molecule, which is either double-stranded or single-stranded, and which comprise at least one

polynucleotide of interest that is sought to be transferred in a cell host or in a unicellular or multicellular host organism.

The term "primer" denotes a specific oligonucleotide sequence which is complementary to a target nucleotide sequence and used to hybridize to the target nucleotide sequence. A primer serves as an initiation point for nucleotide polymerization catalyzed by either DNA polymerase, RNA polymerase or reverse transcriptase.

The term "probe" denotes a defined nucleic acid segment (or nucleotide analog segment, e.g., polynucleotide as defined hereinbelow) which can be used to identify a specific polynucleotide sequence present in samples, said nucleic acid segment comprising a nucleotide sequence complementary of the specific polynucleotide sequence to be identified.

The terms "trait" and "phenotype" are used interchangeably herein and refer to any visible, detectable or otherwise measurable property of an organism such as symptoms of, or susceptibility to a disease for example.

The term "allele" is used herein to refer to variants of a nucleotide sequence. A biallelic polymorphism has two forms. Diploid organisms may be homozygous or heterozygous for an allelic form.

The term "genotype" as used herein refers the identity of the alleles present in an individual or a sample. In the context of the present invention, a genotype preferably refers to the description of the biallelic marker alleles present in an individual or a sample. The term "genotyping" a sample or an individual for a biallelic marker involves determining the specific allele or the specific nucleotide carried by an individual at a biallelic marker.

The term "mutation" as used herein refers to a difference in DNA sequence between or among different genomes or individuals which has a frequency below 1%.

The term "polymorphism" as used herein refers to the occurrence of two or more alternative genomic sequences or alleles between or among different genomes or individuals. "Polymorphic" refers to the condition in which two or more variants of a specific genomic sequence can be found in a population. A "polymorphic site" is the locus at which the variation occurs. A single nucleotide polymorphism is the replacement of one nucleotide by another nucleotide at the polymorphic site. Deletion of a single nucleotide or insertion of a single nucleotide also gives rise to single nucleotide polymorphisms. In the context of the present invention, "single nucleotide polymorphism" preferably refers to a single nucleotide substitution. Typically, between different individuals, the polymorphic site may be occupied by two different nucleotides.

The term "biallelic polymorphism" and "biallelic marker" are used interchangeably herein to refer to a single nucleotide polymorphism having two alleles at a fairly high frequency in the population. A "biallelic marker allele" refers to the nucleotide variants present at a biallelic marker site.

The location of nucleotides in a polynucleotide with respect to the center of the polynucleotide are described herein in the following manner. When a polynucleotide has an odd number of nucleotides, the nucleotide at an equal distance from the 3' and 5' ends of the polynucleotide is considered to be "at the center" of the polynucleotide, and any nucleotide immediately adjacent to the nucleotide at the center, or the nucleotide at the center itself is considered to be "within 1 nucleotide of the center." With an odd number of nucleotides in a polynucleotide any of the five nucleotides positions in the middle of the polynucleotide would be considered to be within 2 nucleotides of the center, and so on. When a polynucleotide has an even number of nucleotides, there would be a bond and not a nucleotide at the center of the polynucleotide. Thus, either of the two central nucleotides would be considered to be "within 1 nucleotide of the center" and any of the four nucleotides in the middle of the polynucleotide would be considered to be "within 2 nucleotides of the center", and so on.

Biallelic markers can be defined as genome-derived polynucleotides having between 2 and 100, preferably between 20, 30, or 40 and 60, and more preferably about 47 nucleotides in length, which exhibit biallelic polymorphism at one single base position. Each biallelic marker therefore corresponds to two forms of a polynucleotide sequence included in a gene which, when compared with one another, present a nucleotide modification at one position.

The term "upstream" is used herein to refer to a location which is toward the 5' end of the polynucleotide from a specific reference point.

The terms "base paired" and "Watson & Crick base paired" are used interchangeably herein to refer to nucleotides which can be hydrogen bonded to one another be virtue of their sequence identities in a manner like that found in double-helical DNA with thymine or uracil residues linked to adenine residues by two hydrogen bonds and cytosine and guanine residues linked by three hydrogen bonds (See Stryer, L., *Biochemistry*, 4th edition, 1995).

The terms "complementary" or "complement thereof" are used herein to refer to the sequences of polynucleotides which is capable of forming Watson & Crick base pairing with another specified polynucleotide throughout the entirety of the complementary region. For the purpose of the present invention, a first polynucleotide is deemed to be complementary to a second polynucleotide when each base in the first polynucleotide is paired with its complementary base. Complementary bases are, generally, A and T (or A and U), or C and G. "Complement" is used herein as a synonym from "complementary polynucleotide", "complementary nucleic acid" and "complementary nucleotide sequence". These terms are applied to pairs of polynucleotides based solely upon their sequences and not any particular set of conditions under which the two polynucleotides would actually bind.

Variants and fragments

1- Polynucleotides

The invention also relates to variants and fragments of the polynucleotides described herein, particularly of an olfactory receptor gene containing one or more biallelic markers according to the invention.

Variants of polynucleotides, as the term is used herein, are polynucleotides that differ from a reference polynucleotide. A variant of a polynucleotide may be a naturally occurring variant such as a naturally occurring allelic variant, or it may be a variant that is not known to occur naturally. Such non-naturally occurring variants of the polynucleotide may be made by mutagenesis techniques, including those applied to polynucleotides, cells or organisms. Generally, differences are limited so that the nucleotide sequences of the reference and the variant are closely similar overall and, in many regions, identical.

Variants of polynucleotides according to the invention include, without being limited to, nucleotide sequences at least 95% identical to a nucleic acid selected from the group consisting of SEQ ID Nos 1-11, or to any polynucleotide fragment of at least 12 consecutive nucleotides from a nucleic acid selected from the group consisting of SEQ ID Nos 1-11, and preferably at least 99% identical, more particularly at least 99.5% identical, and most preferably at least 99.8% identical to a nucleic acid selected from the group consisting of SEQ ID Nos 1-11, or to any polynucleotide fragment of at least 12 consecutive nucleotides from a nucleic acid selected from the group consisting of SEQ ID Nos 1-11.

Changes in the nucleotide of a variant may be silent, which means that they do not alter the amino acids encoded by the polynucleotide. However, nucleotide changes may also result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence. The substitutions, deletions or additions may involve one or more nucleotides. The variants may be altered in coding or non-coding regions or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or additions.

In the context of the present invention, particularly preferred embodiments are those in which the polynucleotides encode polypeptides which retain substantially the same biological function or activity as the mature olfactory receptor protein, or those in which the polynucleotides encode polypeptides which maintain or increase a particular biological activity, while reducing a second biological activity.

A polynucleotide fragment is a polynucleotide which sequence is fully comprised within part of a given nucleotide sequence, preferably the nucleotide sequence of an olfactory receptor gene of the invention, and variants thereof. The fragment can be a portion of a coding or non-coding region of the olfactory receptor gene cluster. Preferably, such fragments comprise at least one of the biallelic markers A1 to A13 or the complements thereto or a biallelic marker in linkage

disequilibrium with one or more of the biallelic markers A1 to A13, for which the respective locations in the sequence listing are provided in Table 2.

Such fragments may be "free-standing", i.e. not part of or fused to other polynucleotides, or they may be comprised within a single larger polynucleotide of which they form a part or region. However, several fragments may be comprised within a single larger polynucleotide.

As representative examples of polynucleotide fragments of the invention, there may be mentioned those which have from about 4, 6, 8, 15, 20, 25, 40, 10 to 30, 30 to 55, 50 to 100, 75 to 100 or 100 to 200 nucleotides in length. Preferred are those fragments having about 47 nucleotides in length, such as those comprising at least one of the biallelic markers A1 to A13 of the olfactory receptor gene. Optionally, such fragments may consist of, or consist essentially of a contiguous span of at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 70, 80, 100, 250, 500 or 1000 nucleotides in length. A set of preferred fragments contain at least one of the biallelic markers A1 to A13 of the olfactory receptor gene which are described herein or the complements thereto.

2- Polypeptides

The invention also relates to variants, fragments, analogs and derivatives of the polypeptides described herein, including mutated olfactory receptor proteins.

The variant may be 1) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue and such substituted amino acid residue may or may not be one encoded by the genetic code, or 2) one in which one or more of the amino acid residues includes a substituent group, or 3) one in which the mutated olfactory receptor is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or 4) one in which the additional amino acids are fused to the mutated olfactory receptor, such as a leader or secretory sequence or a sequence which is employed for purification of the mutated olfactory receptor or a preprotein sequence. Such variants are deemed to be within the scope of those skilled in the art.

In the case of an amino acid substitution in the amino acid sequence of a polypeptide according to the invention, one or several amino acids can be replaced by "equivalent" amino acids. The expression "equivalent" amino acid is used herein to designate any amino acid that may be substituted for one of the amino acids having similar properties, such that one skilled in the art of peptide chemistry would expect the secondary structure and hydropathic nature of the polypeptide to be substantially unchanged. Generally, the following groups of amino acids represent equivalent changes: (1) Ala, Pro, Gly, Glu, Asp, Gln, Asn, Ser, Thr; (2) Cys, Ser, Tyr, Thr; (3) Val, Ile, Leu, Met, Ala, Phe; (4) Lys, Arg, His; (5) Phe, Tyr, Trp, His.

More particularly, a variant olfactory receptor polypeptide comprises amino acid changes ranging from 1, 2, 3, 4, 5, 10 to 20 substitutions, additions or deletions of one aminoacid, preferably from 1 to 10, more preferably from 1 to 5 and most preferably from 1 to 3 substitutions, additi ns or deletions of one amino acid. The preferred amino acid changes are those which have little or no

influence on the biological activity or the capacity of the variant olfactory receptor polypeptide to bind to antibodies raised against a native olfactory receptor protein.

A specific, but not restrictive, embodiment of a modified peptide molecule of interest according to the present invention, which consists in a peptide molecule which is resistant to proteolysis, is a peptide in which the -CONH- peptide bond is modified and replaced by a (CH_2NH) reduced bond, a (NHCO) retro inverso bond, a (CH_2-O) methylene-oxy bond, a (CH_2-S) thiomethylene bond, a (CH_2CH_2) carba bond, a (CO-CH_2) cetomethylene bond, a (CHOH-CH_2) hydroxyethylene bond), a (N-N) bound, a E-alcene bond or also a -CH=CH- bond.

The polypeptide according to the invention could have post-translational modifications. For example, it can present the following modifications: acylation, disulfide bond formation, prenylation, carboxymethylation and phosphorylation.

A polypeptide fragment is a polypeptide which sequence is fully comprised within part of a given polypeptide sequence, preferably a polypeptide encoded by an olfactory receptor gene and variants thereof.

Such fragments may be "free-standing", i.e. not part of or fused to other polypeptides, or they may be comprised within a single larger polypeptide of which they form a part or region. However, several fragments may be comprised within a single larger polypeptide.

As representative examples of polypeptide fragments of the invention, there may be mentioned those which have from about 5, 6, 7, 8, 9 or 10 to 15, 10 to 20, 15 to 40, or 30 to 55 amino acids long. Preferred polypeptide fragments according to the invention comprise a contiguous span of at least 6 amino acids, preferably at least 8 or amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of one amino acid sequence. Preferred are those fragments containing at least one amino acid mutation in the olfactory receptor protein under consideration.

Identity between nucleic acids or polypeptides

The terms "percentage of sequence identity" and "percentage homology" are used interchangeably herein to refer to comparisons among polynucleotides and polypeptides, and are determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. Homology is evaluated using either any of the variety of sequence comparison algorithms and programs known in the art, or by eye inspection. Such algorithms and programs include, but are by no means limited to, TBLASTN, BLASTP, FASTA, TFASTA, and CLUSTALW (Pearson and Lipman, 1988; Altschul et al., 1990; Thompson

et al., 1994; Higgins et al., 1996; Altschul et al., 1990; Altschul et al., 1993). In a particularly preferred embodiment, protein and nucleic acid sequence homologies are evaluated using the Basic Local Alignment Search Tool ("BLAST") which is well known in the art (see, e.g., Karlin and Altschul, 1990; Altschul et al., 1990, 1993, 1997). In particular, five specific BLAST programs are used to perform the following task:

(1) BLASTP and BLAST3 compare an amino acid query sequence against a protein sequence database;

(2) BLASTN compares a nucleotide query sequence against a nucleotide sequence database;

(3) BLASTX compares the six-frame conceptual translation products of a query nucleotide sequence (both strands) against a protein sequence database;

(4) TBLASTN compares a query protein sequence against a nucleotide sequence database translated in all six reading frames (both strands); and

(5) TBLASTX compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.

The BLAST programs identify homologous sequences by identifying similar segments, which are referred to herein as "high-scoring segment pairs," between a query amino or nucleic acid sequence and a test sequence which is preferably obtained from a protein or nucleic acid sequence database. High-scoring segment pairs are preferably identified (i.e., aligned) by means of a scoring matrix, many of which are known in the art. Preferably, the scoring matrix used is the BLOSUM62 matrix (Gonnet et al., 1992; Henikoff and Henikoff, 1993). Less preferably, the PAM or PAM250 matrices may also be used (see, e.g., Schwartz and Dayhoff, eds., 1978). The BLAST programs evaluate the statistical significance of all high-scoring segment pairs identified, and preferably selects those segments which satisfy a user-specified threshold of significance, such as a user-specified percent homology. Preferably, the statistical significance of a high-scoring segment pair is evaluated using the statistical significance formula of Karlin (see, e.g., Karlin and Altschul, 1990). The BLAST programs may be used with the default parameters or with modified parameters provided by the user.

Stringent Hybridization Conditions

By way of example and not limitation, procedures using conditions of high stringency are as follows: Prehybridization of filters containing DNA is carried out for 8 h to overnight at 65°C in buffer composed of 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 µg/ml denatured salmon sperm DNA. Filters are hybridized for 48 h at 65°C, the preferred hybridization temperature, in prehybridization mixture containing 100 µg/ml denatured salmon sperm DNA and 5-20 X 10^6 cpm of ^{32}P-labeled probe. Alternatively, the hybridization step can be performed at 65°C in the presence of SSC buffer, 1 x SSC corresponding to 0.15M NaCl and 0.05 M Na citrate. Subsequently, filt r washes can be done at 37°C for 1 h in a solution containing 2

x SSC, 0.01% PVP, 0.01% Ficoll, and 0.01% BSA, followed by a wash in 0.1 X SSC at 50°C for 45 min. Alternatively, filter washes can be performed in a solution containing 2 x SSC and 0.1% SDS, or 0.5 x SSC and 0.1% SDS, or 0.1 x SSC and 0.1% SDS at 68°C for 15 minute intervals. Following the wash steps, the hybridized probes are detectable by autoradiography. Other conditions of high stringency which may be used are well known in the art and as cited in Sambrook et al., 1989; and Ausubel et al., 1989. These hybridization conditions are suitable for a nucleic acid molecule of about 20 nucleotides in length. There is no need to say that the hybridization conditions described above are to be adapted according to the length of the desired nucleic acid, following techniques well known to the one skilled in the art. The suitable hybridization conditions may for example be adapted according to the teachings disclosed in the book of Hames and Higgins (1985) or in Sambrook et al.(1989).

HOMOLOGIES OF THE NOVEL OLFACTORY RECEPTOR GENE WITH KNOWN OLFACTORY RECEPTORS

A comparison analysis of various olfactory receptor amino acid sequences, including the novel sequences of the invention, has been performed with the alignment program Pileup and the translation program MAP (Winsconsin Package version 8, GCG). The protein sequences were sorted into different families and subfamilies, taking into account their Amino acid Sequence Identity (ASI). It was observed the Open Reading Frames of the OLF1 to OLF10 genes are genetically clearly distinguished from the already known olfactory receptor sequences. For example, the olfactory receptor OLF2 presents respectively 39.9 %, 43.1 % and 44.2 % of identity with prior art olfactory receptors referred in Genbank as L35475, U58675_1 and Y10530. In addition, the nucleotide sequences of Orf-2 to Orf-10 according to the invention are all grouped together, whereas the nucleotide Orf-1 of the invention forms a new family by itself. These amino acid sequence comparison data clearly indicate that the novel olfactory receptor sequences of the invention share common genetic characteristics (Orf-2 to Orf-10) or have specific characteristics (Orf-1) that are not found in the prior art olfactory receptor sequences.

A. OLF1 TO OLF10 GENE POLYNUCLEOTIDES.

The cluster of ten olfactory receptor genes has been found by the inventors to be located on the human chromosome 11, more precisely within the 11q12-q13 locus of said chromosome as described in Example 1.

1. Genomic sequences of the olfactory receptor gene

The present invention concerns the genomic sequence of an olfactory receptor cluster. The present invention encompasses the olfactory receptor gene, or olfactory receptor genomic sequences consisting of, consisting essentially of, or comprising the sequence of SEQ ID No 1, a sequence

complementary thereto, as well as fragments and variants thereof. These polynucleotides may be purified, isolated, or recombinant.

The invention also encompasses a purified, isolated, or recombinant polynucleotide comprising a nucleotide sequence having at least 70, 75, 80, 85, 90, or 95% nucleotide identity with a nucleotide sequence of SEQ ID No 1 or a complementary sequence thereto or a fragment thereof. The nucleotide differences as regards to the nucleotide sequence of SEQ ID No 1 may be generally randomly distributed throughout the entire nucleic acid. Nevertheless, preferred nucleic acids are those wherein the nucleotide differences as regards to the nucleotide sequence of SEQ ID No 1 are predominantly located outside the coding sequences contained in the exons. These nucleic acids, as well as their fragments and variants, may be used as oligonucleotide primers or probes in order to detect the presence of a copy of the olfactory receptor gene in a test sample, or alternatively in order to amplify a target nucleotide sequence within the olfactory receptor sequences.

Another object of the invention consists of a purified, isolated, or recombinant nucleic acid that hybridizes with the nucleotide sequence of SEQ ID No 1 or a complementary sequence thereto, under stringent hybridization conditions as defined above.

Particularly preferred nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 1: 1-113643, 114064-127488, 127855-144460. Additional preferred nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 1: 1-10000, 10001-20000, 20001-30000, 30001-40000, 40001-50000, 50001-60000, 60001-70000, 70001-80000, 80001-90000, 90001-100000, 100001-110000, 110001-120000, 120001-130000, 130001-140000, and 140001-144460. Further preferred nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 1: 1-5000, 5001-10000, 10001-15000, 15001-20000, 20001-25000, 25001-30000, 30001-35000, 35001-40000, 40001-45000, 45001-50000, 50001-55000, 55001-60000, 60001-65000, 65001-70000, 70001-75000, 75001-80000, 80001-85000, 85001-90000, 90001-95000, 95001-100000, 100001-105000, 105001-110000, 110001-115000, 115001-120000, 120001-125000, 125001-130000, 130001-135000, 135001-140000, and 140001-144460.

The olfactory receptor genomic nucleic acid comprises 10 open reading frames, each carried by a single exon and encoding a polypeptide designated OLF1 to OLF10. The open reading frames positions of OLF1 to OLF10 in SEQ ID No 1 are given as features in the sequence listing and are also detailed below in Table A.

Two truncated ubiquitin polypeptides Ubi1 and Ubi2, unrelated to olfactory receptor coding sequences, are encoded on the complementary strand of the olfactory receptor gene. The complementary sequence of the Ubi1 ORF is located between the nucleotide in position 114063 and the nucleotide in position 113644 of the nucleotide sequence of SEQ ID No 1. The complementary sequence of the Ubi2 ORF is located between the nucleotide in position 127854 and the nucleotide in position 127489 of the nucleotide sequence of SEQ ID No 1.

Table A

Coding regions			Non-coding regions		
Name	Position in SEQ ID No 1		Name	Position in SEQ ID No 1	
	Beginning	End		Beginning	End
OLF1	2406	2600	NC1	1	2405
OLF2	9711	10658	NC2	2601	9710
OLF3	24851	25369	NC3	10659	24850
OLF4	45714	46661	NC4	25370	45713
OLF5	80198	81115	NC5	46662	80197
OLF6	96291	96902	NC6	81116	96290
OLF7	110758	111564	NC7	96903	110757
OLF8	122525	122887	NC8	111565	122524
OLF9	132454	133389	NC9	122888	132453
OLF10	143398	143577	NC10	133390	143397
			NC11	143578	144460

Thus, the invention embodies purified, isolated, or recombinant polynucleotides comprising a nucleotide sequence selected from the group consisting of the 10 open reading frames of the olfactory receptor gene, or a sequence complementary thereto.

The nucleic acid of SEQ ID No 1 also comprises non coding portions flanking each of the ten olfactory receptor open reading frames of the sense DNA strand.

The invention also embodies purified, isolated, or recombinant polynucleotides comprising a nucleotide sequence selected from the group consisting of the non-coding regions contained in the olfactory receptor gene cluster of SEQ ID No 1, or a sequence complementary thereto as well as their fragments or variants. The term "non-coding" sequence refers to any nucleotide sequence which does not encode an amino acid. The non-coding sequences encompass upstream and downstream regions of the olfactory receptor ORFs of the invention, as well as regions located

between two successive olfactory receptor ORFs, as indicated in Table A which lists the 11 non-coding regions named from NC1 to NC11.

The nucleic acids defining the non-coding sequences of the polynucleotide of SEQ ID No 1 described above, as well as their fragments and variants, may be used as oligonucleotide primers or probes in order to detect the presence of a copy of one of the olfactory receptor genes of the invention in a test sample, or alternatively in order to amplify a target nucleotide sequence within the cluster of olfactory receptor encoding sequences according to the invention.

While this section is entitled "Genomic Sequences of the olfactory receptor gene," it should be noted that nucleic acid fragments of any size and sequence may also be comprised by the polynucleotides described in this section, flanking the genomic sequences of olfactory receptor on either side or between two or more such genomic sequences.

2. Coding regions of the olfactory receptor gene

The 10 olfactory receptor open reading frames are presented individually as SEQ ID Nos 2-11 in the appended sequence listing.

Thus, another object of the invention is a purified, isolated, or recombinant nucleic acid comprising a nucleotide sequence selected from the group consisting of SEQ ID Nos 2-11, complementary sequences thereto, as well as allelic variants, and fragments thereof. Moreover, preferred polynucleotides of the invention include purified, isolated, or recombinant olfactory receptor cDNAs consisting of, consisting essentially of, or comprising a sequence selected from the group consisting of SEQ ID Nos 2-11.

The invention also pertains to a purified or isolated nucleic acid comprising a polynucleotide having at least 95% nucleotide identity with a polynucleotide selected from the group consisting of SEQ ID Nos 2-11, advantageously 99 % nucleotide identity, preferably 99.5% nucleotide identity and most preferably 99.8% nucleotide identity with a polynucleotide selected from the group consisting of SEQ ID Nos 2-11, or a sequence complementary thereto or a biologically active fragment thereof.

Another object of the invention relates to purified, isolated or recombinant nucleic acids comprising a polynucleotide that hybridizes, under the stringent hybridization conditions defined herein, with a polynucleotide selected from the group consisting of SEQ ID Nos 2-11, or a sequence complementary thereto or a biologically active fragment thereof.

Particularly preferred nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of a sequence selected from the group consisting of SEQ ID Nos 2-11 or the complements thereof. Additional preferred embodiments of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span f at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of a sequence selected from the group consisting of SEQ ID Nos 2-11 r the

complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of said selected sequence : 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-450, 451-500, 501-550, 551-600, 601-650, 651-700, 701-750, 751-800, 801-850, 851-900, 901- the terminal nucleotide of the olfactory receptor coding regions, to the extent that such nucleotide positions are consistent with the lengths of the particular olfactory receptor coding region being referred to. Further preferred embodiments of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of a sequence selected from the group consisting of SEQ ID Nos 2, 4, 7, 9 and 11, or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of said selected sequence: 1-25, 26-50, 51-75, 76-100, 101-125, 126-150, 151-175, 176-200, 201-225, 226-250, 251-275, 276-300, 301-325, 326-350, 351-375, 376-400, 401-425, 426-450, 451-475, 476-500, 501-525, 526-550, 551-575, 576-the terminal nucleotide of the olfactory receptor coding regions, to the extent that such nucleotide positions are consistent with the lengths of the particular olfactory receptor coding region being referred to.

The present invention also embodies isolated, purified, and recombinant polynucleotides encoding olfactory receptor polypeptides, wherein olfactory receptor polypeptides comprise an amino acid sequence selected from the group consisting of SEQ ID Nos 12-21, a nucleotide sequence complementary thereto, a fragment or a variant thereof. The present invention also embodies isolated, purified, and recombinant polynucleotides which encode polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a sequence selected from the group consisting of SEQ ID Nos 12-21. In a preferred embodiment, the present invention embodies isolated, purified, and recombinant polynucleotides which encode polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a sequence selected from the group consisting of SEQ ID Nos 12-21 wherein said contiguous span includes at least 1, 2, 3, 5 or 10 of the following amino acid positions in said selected sequence: 1-20, 21-40, 41-60, 61-80, 81-100, 101-120, 121-140, 141-160, 161-180, 181-200, 201-220, 221-240, 241-260, 261-280, 281-300, 301-the terminal amino acid of the olfactory receptor proteins, to the extent that such amino acid positions are consistent with the lengths of the particular olfactory receptor protein being referred to. In another preferred embodiment, the present invention embodies isolated, purified, and recombinant polynucleotides which encode polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amin acids of a sequence selected from the group consisting of SEQ ID Nos 12, 14, 17, 19 or 21 wherein said contiguous span includes at least 1, 2, 3, 5 or 7 of the following amino acid positions in said selected sequence: 1-10, 11-20, 21-30, 31-40, 41-50, 51-60, 61-70, 71-80, 81-90, 91-100, 101-

110, 111-120, 121-130, 131-140, 141-150, 151-160, 161-170, 171-180, 181-190, 191-the terminal amino acid of the olfactory receptor proteins, to the extent that such amino acid positions are consistent with the lengths of the particular olfactory receptor protein being referred to.

In further preferred embodiments, the present invention embodies isolated, purified, and recombinant polynucleotides which encode olfactory receptor polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a sequence selected from the group consisting of SEQ ID No 12-21, wherein said contiguous span includes at least one amino acid at the following positions of said selected sequence:

i) 1-3, 10, 16, 21, 28, 33, 34, 36, 42-44, 46, 49, 53, 54, 57, 59, 63, and 64 for SEQ ID No 12;

ii) 2, 4, 6, 8, 18, 25, 34, 37, 44, 52, 56, 80, 83, 89, 98, 101, 102, 113, 114, 117, 120, 139, 148, 158, 186, 195, 212, 219, 247, 266, 270, 280, 295, 298, 299, 301, 311, and 313-315 for SEQ ID No 13;

iii) 2-4, 6, 18, 21, 25, 34, 37, 98, 99, 102, 113, 114, 133, 143, 148, 158-163, 166, 167, 169, and 170 for SEQ ID No 14;

iv) 2, 4, 6, 8, 18, 25, 34, 37, 44, 52, 54, 56, 80, 83, 89, 98, 101, 102, 113, 114, 117, 120, 139, 148, 158, 186, 195, 212, 219, 247, 266, 270, 280, 298, 299, 311, and 313-315 for SEQ ID No 15;

v) 3, 18, 20, 25, 34, 47, 49, 67, 97, 100, 107, 108, 112, 113, 126, 135, 142, 146, 147, 157, 159-160, 194, 196, 228, 245, 264, 265, 269, 279, 298, and 302 for SEQ ID No 16;

vi) 2, 6, 18, 20, 33, 34, 37, 65, 68, 69, 72, 86, 88, 101, 107, 113, 114, 148, 158, 161, 164, 195, and 198 for SEQ ID No 17;

vii) 2, 6, 7, 52, 56, 67, 88, 94, 97, 110, 113, 116, 119, 120, 127, 135, 150, 153, 164, 174, 175, 180, 184, 217, 221, 259, 261, and 268 for SEQ ID No 18;

viii) 17, 18, 20, 28, 33, 35, 49-52, 105, 111, and 112 for SEQ ID No 19;

ix) 17, 20, 33, 35, 49-53, 56, 111, 112, 132, 138, 141, 147, 154, 157, 160, 163, 164, 194, 197, 204, 211, 214, 218, 219, 252, 265, 286, 295, 301, 303, 305, 306 and 309 for SEQ ID No 20; and

x) 9, 18, 26-28, 34, 47 and 50 for SEQ ID No 21, to the extent that such amino acid lengths are consistent with the lengths of the particular olfactory receptor protein being referred to.

Additional preferred fragments of the nucleotide sequences of SEQ ID Nos 2-11 are those encoding olfactory receptor polypeptide fragments located outside the transmembrane domains of the corresponding protein as located in boxes in Figure 1.

The above disclosed polynucleotides that contain only coding sequences derived from the olfactory receptor ORFs may be expressed in a desired host cell or a desired host organism, when said polynucleotides are placed under the control of suitable expression signals. Such a polynucleotide, when placed under suitable expression signals, may be inserted in a vector for its expression.

While this section is entitled " Coding regions of the olfactory receptor gene," it should be noted that nucleic acid fragments of any size and sequence may also be comprised by the polynucleotides described in this section, flanking the genomic sequences of olfactory receptor on either side or between two or more such genomic sequences.

3. Polynucleotide Constructs

The terms "polynucleotide construct" and "recombinant polynucleotide" are used interchangeably herein to refer to linear or circular, purified or isolated polynucleotides that have been artificially designed and which comprise at least two nucleotide sequences that are not found as contiguous nucleotide sequences in their initial natural environment.

DNA Construct That Enables Directing Temporal And Spatial olfactory receptor Gene Expression In Recombinant Cell Hosts And In Transgenic Animals.

In order to study the physiological and phenotypic consequences of a lack of synthesis of the olfactory receptor protein, both at the cell level and at the multi cellular organism level, the invention also encompasses DNA constructs and recombinant vectors enabling a conditional expression of a specific allele of the olfactory receptor genomic sequence or cDNA and also of a copy of this genomic sequence or cDNA harboring substitutions, deletions, or additions of one or more bases as regards to the olfactory receptor nucleotide sequence of SEQ ID Nos 1-11, or a fragment thereof, these base substitutions, deletions or additions being located in the coding regions of the olfactory receptor genomic sequence or within the olfactory receptor open reading frames of SEQ ID Nos 2-11. In a preferred embodiment, the olfactory receptor sequence comprises a biallelic marker of the present invention. In a preferred embodiment, the olfactory receptor sequence comprises a biallelic marker of the present invention, preferably one of the biallelic markers A1 to A13.

The present invention embodies recombinant vectors comprising any one of the polynucleotides described in the present invention. More particularly, the polynucleotide constructs according to the present invention can comprise any of the polynucleotides described in the "Genomic sequences of the olfactory receptor gene" section, the "Coding regions of the olfactory receptor Gene" section, and the "Oligonucleotide probes and primers" section.

DNA Constructs Allowing Homologous Recombination: Replacement Vectors

A first preferred DNA construct will comprise, from 5'-end to 3'-end: (a) a first nucleotide sequence that is comprised in the olfactory receptor genomic sequence; (b) a nucleotide sequence comprising a positive selection marker, such as the marker for neomycine resistance (*neo*); and (c) a second nucleotide sequence that is comprised in the olfactory receptor genomic sequence, and is located on the genome downstream the first olfactory receptor nucleotide sequence (a).

In a preferred embodiment, this DNA construct also comprises a negative selection marker located upstream the nucleotide sequence (a) or downstream the nucleotide sequence (c). Preferably, the negative selection marker comprises the thymidine kinase (*tk*) gene (Thomas et al., 1986), the hygromycine beta gene (Te Riele et al., 1990), the *hprt* gene (Van der Lugt et al., 1991; Reid et al., 1990) or the Diphteria toxin A fragment (*Dt-A*) gene (Nada et al., 1993; Yagi et al.1990). Preferably, the positive selection marker is located within an olfactory receptor open reading frame sequence so as to interrupt the sequence encoding an olfactory receptor protein. These replacement vectors are described, for example, by Thomas et al.(1986; 1987), Mansour et al.(1988) and Koller et al.(1992).

The first and second nucleotide sequences (a) and (c) may be indifferently located within an olfactory receptor regulatory sequence, an intronic sequence, an exon sequence or a sequence containing both regulatory and/or intronic and/or exon sequences. The size of the nucleotide sequences (a) and (c) ranges from 1 to 50 kb, preferably from 1 to 10 kb, more preferably from 2 to 6 kb and most preferably from 2 to 4 kb.

DNA Constructs Allowing Homologous Recombination: Cre-LoxP System.

These new DNA constructs make use of the site specific recombination system of the P1 phage. The P1 phage possesses a recombinase called Cre which interacts specifically with a 34 base pairs *lox*P site. The *lox*P site is composed of two palindromic sequences of 13 bp separated by a 8 bp conserved sequence (Hoess et al., 1986). The recombination by the Cre enzyme between two *lox*P sites having an identical orientation leads to the deletion of the DNA fragment.

The Cre-*lox*P system used in combination with a homologous recombination technique has been first described by Gu et al.(1993, 1994). Briefly, a nucleotide sequence of interest to be inserted in a targeted location of the genome harbors at least two *lox*P sites in the same orientation and located at the respective ends of a nucleotide sequence to be excised from the recombinant genome. The excision event requires the presence of the recombinase (Cre) enzyme within the nucleus of the recombinant cell host. The recombinase enzyme may be brought at the desired time either by (a) incubating the recombinant cell hosts in a culture medium containing this enzyme, by injecting the Cre enzyme directly into the desired cell, such as described by Araki et al.(1995), or by lipofection of the enzyme into the cells, such as described by Baubonis et al.(1993); (b) transfecting the cell host with a vector comprising the *Cre* coding sequence operably linked to a promoter functional in the recombinant cell host, which promoter being optionally inducible, said vector being

introduced in the recombinant cell host, such as described by Gu et al.(1993) and Sauer et al.(1988); (c) introducing in the genome of the cell host a polynucleotide comprising the *Cre* coding sequence operably linked to a promoter functional in the recombinant cell host, which promoter is optionally inducible, and said polynucleotide being inserted in the genome of the cell host either by a random insertion event or an homologous recombination event, such as described by Gu et al.(1994).

In a specific embodiment, the vector containing the sequence to be inserted in the olfactory receptor gene by homologous recombination is constructed in such a way that selectable markers are flanked by *lox*P sites of the same orientation, it is possible, by treatment by the Cre enzyme, to eliminate the selectable markers while leaving the olfactory receptor sequences of interest that have been inserted by an homologous recombination event. Again, two selectable markers are needed: a positive selection marker to select for the recombination event and a negative selection marker to select for the homologous recombination event. Vectors and methods using the Cre-*lox*P system are described by Zou et al.(1994).

Thus, a second preferred DNA construct of the invention comprises, from 5'-end to 3'-end: (a) a first nucleotide sequence that is comprised in the olfactory receptor genomic sequence; (b) a nucleotide sequence comprising a polynucleotide encoding a positive selection marker, said nucleotide sequence comprising additionally two sequences defining a site recognized by a recombinase, such as a *lox*P site, the two sites being placed in the same orientation; and (c) a second nucleotide sequence that is comprised in the olfactory receptor genomic sequence, and is located on the genome downstream of the first olfactory receptor nucleotide sequence (a).

The sequences defining a site recognized by a recombinase, such as a *lox*P site, are preferably located within the nucleotide sequence (b) at suitable locations bordering the nucleotide sequence for which the conditional excision is sought. In one specific embodiment, two *lox*P sites are located at each side of the positive selection marker sequence, in order to allow its excision at a desired time after the occurrence of the homologous recombination event.

In a preferred embodiment of a method using the third DNA construct described above, the excision of the polynucleotide fragment bordered by the two sites recognized by a recombinase, preferably two loxP sites, is performed at a desired time, due to the presence within the genome of the recombinant host cell of a sequence encoding the Cre enzyme operably linked to a promoter sequence, preferably an inducible promoter, more preferably a tissue-specific promoter sequence and most preferably a promoter sequence which is both inducible and tissue-specific, such as described by Gu et al.(1994).

The presence of the Cre enzyme within the genome of the recombinant cell host may result from the breeding of two transgenic animals, the first transgenic animal bearing the olfactory receptor-derived sequence of interest containing the *lox*P sites as described above and the second transgenic animal bearing the *Cre* coding sequence operably linked to a suitable promoter sequence, such as described by Gu et al.(1994).

Spatio-temporal control of the Cre enzyme expression may also be achieved with an adenovirus based vector that contains the Cre gene thus allowing infection of cells, or *in vivo* infection of organs, for delivery of the Cre enzyme, such as described by Anton and Graham (1995) and Kanegae et al.(1995).

The DNA constructs described above may be used to introduce a desired nucleotide sequence of the invention, preferably an olfactory receptor genomic sequence or an olfactory receptor coding region sequences, and most preferably an altered copy of an olfactory receptor genomic or coding region sequences, within a predetermined location of the targeted genome, leading either to the generation of an altered copy of a targeted gene (knock-out homologous recombination) or to the replacement of a copy of the targeted gene by another copy sufficiently homologous to allow an homologous recombination event to occur (knock-in homologous recombination). In a specific embodiment, the DNA constructs described above may be used to introduce an olfactory receptor genomic sequence or an olfactory receptor coding region sequence comprising at least one biallelic marker of the present invention, preferably at least one biallelic marker selected from the group consisting of A1 to A13.

Nuclear Antisense DNA Constructs

Other compositions containing a vector of the invention comprising an oligonucleotide fragment of the nucleic sequence SEQ ID Nos 2-11, preferably a fragment including the start codon of the olfactory receptor gene, as an antisense tool that inhibits the expression of the corresponding olfactory receptor gene. Preferred methods using antisense polynucleotide according to the present invention are the procedures described by Sczakiel et al.(1995) or those described in PCT Application No WO 95/24223.

Preferred antisense polynucleotides according to the present invention are complementary to a sequence of the mRNAs of olfactory receptor that contains the translation initiation codon ATG.

Preferably, the antisense polynucleotides of the invention have a 3' polyadenylation signal that has been replaced with a self-cleaving ribozyme sequence, such that RNA polymerase II transcripts are produced without poly(A) at their 3' ends, these antisense polynucleotides being incapable of export from the nucleus, such as described by Liu et al.(1994). In a preferred embodiment, these olfactory receptor antisense polynucleotides also comprise, within the ribozyme cassette, a histone stem-loop structure to stabilize cleaved transcripts against 3'-5' exonucleolytic degradation, such as the structure described by Eckner et al.(1991).

4. Oligonucleotide probes and primers

Polynucleotides derived from the olfactory receptor gene are useful in order to detect the presence of at least a copy of a nucleotide sequence of SEQ ID Nos 1-11, or a fragment, complement, or variant thereof in a test sample, preferably a human olfactory epithelium tissue or isolated human olfactory epithelium cells.

Particularly preferred probes and primers of the invention include isolated, purified, or recombinant polynucleotides c mprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 1: 1-113643, 114064-127488, 127855-144460. Additional preferred probes and primers of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 1: 1-10000, 10001-20000, 20001-30000, 30001-40000, 40001-50000, 50001-60000, 60001-70000, 70001-80000, 80001-90000, 90001-100000, 100001-110000, 110001-120000, 120001-130000, 130001-140000, and 140001-144460. Further preferred probes and primers of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 1: 1-5000, 5001-10000, 10001-15000, 15001-20000, 20001-25000, 25001-30000, 30001-35000, 35001-40000, 40001-45000, 45001-50000, 50001-55000, 55001-60000, 60001-65000, 65001-70000, 70001-75000, 75001-80000, 80001-85000, 85001-90000, 90001-95000, 95001-100000, 100001-105000, 105001-110000, 110001-115000, 115001-120000, 120001-125000, 125001-130000, 130001-135000, 135001-140000, and 140001-144460.

Other particularly preferred probes and primers of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 45 or 50 nucleotides of a sequence selected from the group consisting of SEQ ID Nos 2-11 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of said selected sequence : 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-450, 451-500, 501-550, 551-600, 601-650, 651-700, 701-750, 751-800, 801-850, 851-900, 901- the terminal nucleotide of the olfactory receptor coding regions, to the extent that such nucleotide positions are consistent with the lengths of the particular olfactory receptor coding region being referred to. Further preferred probes and primers of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 22 or 25 nucleotides of a sequence selected from the group consisting of SEQ ID Nos 2, 4, 7, 9 and 11, or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of said selected sequence: 1-25, 26-50, 51-75, 76-100, 101-125, 126-150, 151-175, 176-200, 201-225, 226-250, 251-275, 276-300, 301-325, 326-350, 351-375, 376-400, 401-425, 426-450, 451-475, 476-500, 501--525, 526-550, 551-575, 576-the terminal nucleotide of the olfactory receptor coding regions, to the extent that such nucleotide

positions are consistent with the lengths of the particular olfactory receptor coding region being referred to.

Thus, the invention also relates to nucleic acid probes characterized in that they hybridize specifically, under the stringent hybridization conditions defined above, with a nucleic acid selected from the group consisting of SEQ ID Nos 1-11, a variant thereof and a sequence complementary thereto.

In one embodiment the invention encompasses isolated, purified, and recombinant polynucleotides consisting of, or consisting essentially of a contiguous span of 8 to 50 nucleotides of SEQ ID No 1 and the complement thereof, wherein said span includes an olfactory receptor-related biallelic marker in said sequence; optionally, wherein said olfactory receptor-related biallelic marker is selected from the group consisting of A1 to A13, and the complements thereof; optionally, wherein said contiguous span is 18 to 47 nucleotides in length and said biallelic marker is within 4 nucleotides of the center of said polynucleotide; optionally, wherein said polynucleotide consists of said contiguous span and said contiguous span is 25 nucleotides in length and said biallelic marker is at the center of said polynucleotide; optionally, wherein the 3' end of said contiguous span is present at the 3' end of said polynucleotide; and optionally, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide and said biallelic marker is present at the 3' end of said polynucleotide. In a preferred embodiment, said probes comprises, consists of, or consists essentially of a sequence selected from the following sequences: P1 to P13 and the complementary sequences thereto, for which the respective locations in the sequence listing are provided in Table 3.

In another embodiment the invention encompasses isolated, purified and recombinant polynucleotides comprising, consisting of, or consisting essentially of a contiguous span of 8 to 50 nucleotides of SEQ ID No 1, or the complements thereof, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide, and wherein the 3' end of said polynucleotide is located within 20 nucleotides upstream of an olfactory receptor-related biallelic marker in said sequence; optionally, wherein said olfactory receptor-related biallelic marker is selected from the group consisting of A1 to A13, and the complements thereof; optionally, wherein the 3' end of said polynucleotide is located 1 nucleotide upstream of said olfactory receptor-related biallelic marker in said sequence; and optionally, wherein said polynucleotide consists essentially of a sequence selected from the following sequences: D1 to D13 and E1 to E13, for which the respective locations in the sequence listing are provided in Table 4.

In a further embodiment, the invention encompasses isolated, purified, or recombinant polynucleotides comprising, consisting of, or consisting essentially of a sequence selected from the following sequences: B1 to B11 and C1 to C11, for which the respective locations in the sequence listing are provided in Table 1.

In an additional embodiment, the invention encompasses polynucleotides for use in hybridization assays, sequencing assays, and enzyme-based mismatch detection assays for

determining the identity of the nucleotide at an olfactory receptor-related biallelic marker in SEQ ID No 1, or the complements thereof, as well as polynucleotides for use in amplifying segments of nucleotides comprising an olfactory receptor-related biallelic marker in SEQ ID No 1, or the complements thereof; optionally, wherein said olfactory receptor-related biallelic marker is selected from the group consisting of A1 to A13, and the complements thereof.

A probe or a primer according to the invention has between 8 and 1000 nucleotides in length, or is specified to be at least 12, 15, 18, 20, 25, 35, 40, 50, 60, 70, 80, 100, 250, 500 or 1000 nucleotides in length. More particularly, the length of these probes and primers can range from 8, 10, 15, 20, or 30 to 100 nucleotides, preferably from 10 to 50, more preferably from 15 to 30 nucleotides. Shorter probes and primers tend to lack specificity for a target nucleic acid sequence and generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. Longer probes and primers are expensive to produce and can sometimes self-hybridize to form hairpin structures. The appropriate length for primers and probes under a particular set of assay conditions may be empirically determined by one of skill in the art. A preferred probe or primer consists of a nucleic acid comprising a polynucleotide selected from the group of the nucleotide sequences of P1 to P13 and the complementary sequence thereto, B1 to B11, C1 to C11, D1 to D13, and E1 to E13.

Primers and other oligonucleotides according to the invention are synthesized to be "substantially" complementary to a strand of the olfactory receptor gene of the invention to be amplified. The primer sequence does not need to reflect the exact sequence of the DNA template. Minor mismatches can be accommodated by reducing the stringency of the hybridization conditions. Among the various methods available to design useful primers, the OSP computer software can be used by the skilled person (see Hillier & Green, 1991). All primers contained a common upstream oligonucleotide tail enabling the easy systematic sequencing of the resulting amplification fragments.

The formation of stable hybrids depends on the melting temperature (Tm) of the DNA. The Tm depends on the length of the primer or probe, the ionic strength of the solution and the G+C content. The higher the G+C content of the primer or probe, the higher is the melting temperature because G:C pairs are held by three H bonds whereas A:T pairs have only two. The GC content in the probes of the invention usually ranges between 10 and 75 %, preferably between 35 and 60 %, and more preferably between 40 and 55 %.

The primers and probes can be prepared by any suitable method, including, for example, cloning and restriction of appropriate sequences and direct chemical synthesis by a method such as the phosphodiester method of Narang et al.(1979), the phosphodiester method of Brown et al.(1979), the diethylphosphoramidite method of Beaucage et al.(1981) and the solid support method described in EP 0 707 592.

Detection probes are generally nucleic acid sequences or uncharged nucleic acid analogs such as, for example peptide nucleic acids which are disclosed in International Patent Application WO 92/20702, morpholino analogs which are described in U.S. Patents Numbered 5,185,444; 5,034,506 and 5,142,047. The probe may have to be rendered "non-extendable" in that additional dNTPs cannot be added to the probe. In and of themselves analogs usually are non-extendable and nucleic acid probes can be rendered non-extendable by modifying the 3' end of the probe such that the hydroxyl group is no longer capable of participating in elongation. For example, the 3' end of the probe can be functionalized with the capture or detection label to thereby consume or otherwise block the hydroxyl group. Alternatively, the 3' hydroxyl group simply can be cleaved, replaced or modified, U.S. Patent Application Serial No. 07/049,061 filed April 19, 1993 describes modifications, which can be used to render a probe non-extendable.

Any of the polynucleotides of the present invention can be labeled, if desired, by incorporating any label known in the art to be detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include radioactive substances (including, ^{32}P, ^{35}S, ^{3}H, ^{125}I), fluorescent dyes (including, 5-bromodesoxyuridin, fluorescein, acetylaminofluorene, digoxigenin) or biotin. Preferably, polynucleotides are labeled at their 3' and 5' ends. Examples of non-radioactive labeling of nucleic acid fragments are described in the French patent No. FR-7810975 or by Urdea et al (1988) or Sanchez-Pescador et al (1988). In addition, the probes according to the present invention may have structural characteristics such that they allow the signal amplification, such structural characteristics being, for example, branched DNA probes as those described by Urdea et al. in 1991 or in the European patent No. EP 0 225 807 (Chiron).

A label can also be used to capture the primer, so as to facilitate the immobilization of either the primer or a primer extension product, such as amplified DNA, on a solid support. A capture label is attached to the primers or probes and can be a specific binding member which forms a binding pair with the solid's phase reagent's specific binding member (e.g. biotin and streptavidin). Therefore depending upon the type of label carried by a polynucleotide or a probe, it may be employed to capture or to detect the target DNA. Further, it will be understood that the polynucleotides, primers or probes provided herein, may, themselves, serve as the capture label. For example, in the case where a solid phase reagent's binding member is a nucleic acid sequence, it may be selected such that it binds a complementary portion of a primer or probe to thereby immobilize the primer or probe to the solid phase. In cases where a polynucleotide probe itself serves as the binding member, those skilled in the art will recognize that the probe will contain a sequence or "tail" that is not complementary to the target. In the case where a polynucleotide primer itself serves as the capture label, at least a portion of the primer will be free to hybridize with a nucleic acid on a solid phase. DNA Labeling techniques are well known to the skilled technician.

The probes of the present invention are useful for a number of purposes. They can be notably used in Southern hybridization to genomic DNA or Northern hybridization to mRNA. The probes can also be used to detect PCR amplification products. They may also be used to detect mismatches in the OLF1 to OLF10 genes or mRNA using other techniques. Generally, the probes are complementary to the OLF1 to OLF10 gene coding sequences, although probes complementary to non-coding sequences are also contemplated. The probes of the present invention can also be useful for genotyping the biallelic markers of the cluster of olfactory receptor genes of the present invention.

Any of the polynucleotides, primers and probes of the present invention can be conveniently immobilized on a solid support. Solid supports are known to those skilled in the art and include the walls of wells of a reaction tray, test tubes, polystyrene beads, magnetic beads, nitrocellulose strips, membranes, microparticles such as latex particles, sheep (or other animal) red blood cells, duracytes and others. The solid support is not critical and can be selected by one skilled in the art. Thus, latex particles, microparticles, magnetic or non-magnetic beads, membranes, plastic tubes, walls of microtiter wells, glass or silicon chips, sheep (or other suitable animal's) red blood cells and duracytes are all suitable examples. Suitable methods for immobilizing nucleic acids on solid phases include ionic, hydrophobic, covalent interactions and the like. A solid support, as used herein, refers to any material which is insoluble, or can be made insoluble by a subsequent reaction. The solid support can be chosen for its intrinsic ability to attract and immobilize the capture reagent. Alternatively, the solid phase can retain an additional receptor which has the ability to attract and immobilize the capture reagent. The additional receptor can include a charged substance that is oppositely charged with respect to the capture reagent itself or to a charged substance conjugated to the capture reagent. As yet another alternative, the receptor molecule can be any specific binding member which is immobilized upon (attached to) the solid support and which has the ability to immobilize the capture reagent through a specific binding reaction. The receptor molecule enables the indirect binding of the capture reagent to a solid support material before the performance of the assay or during the performance of the assay. The solid phase thus can be a plastic, derivatized plastic, magnetic or non-magnetic metal, glass or silicon surface of a test tube, microtiter well, sheet, bead, microparticle, chip, sheep (or other suitable animal's) red blood cells, duracytes® and other configurations known to those of ordinary skill in the art. The polynucleotides of the invention can be attached to or immobilized on a solid support individually or in groups of at least 2, 5, 8, 10, 12, 15, 20, or 25 distinct polynucleotides of the invention to a single solid support. In addition, polynucleotides other than those of the invention may be attached to the same solid support as one or more polynucleotides of the invention.

Consequently, the invention also comprises a method for detecting the presence of a nucleic acid comprising a nucleotide sequence selected from a group consisting of SEQ ID Nos 1-11, a

fragment or a variant thereof and a complementary sequence thereto in a sample. said meth d comprising the following steps of:

a) bringing into contact a nucleic acid probe or a plurality of nucleic acid probes which can hybridize with a nucleotide sequence selected from the group consisting of the nucleotide sequences of SEQ ID Nos 1-11, a fragment or a variant thereof and a complementary sequence thereto and the sample to be assayed; and

b) detecting the hybrid complex formed between the probe and a nucleic acid in the sample.

The invention further concerns a kit for detecting the presence of a nucleic acid comprising a nucleotide sequence selected from a group consisting of SEQ ID Nos 1-11, a fragment or a variant thereof and a complementary sequence thereto in a sample, said kit comprising:

a) a nucleic acid probe or a plurality of nucleic acid probes which can hybridize with a nucleotide sequence selected from the group consisting of the nucleotide sequences of SEQ ID Nos 1-11, a fragment or a variant thereof and a complementary sequence thereto; and

b) optionally, the reagents necessary for performing the hybridization reaction.

In a first preferred embodiment of this detection method and kit, said nucleic acid probe or the plurality of nucleic acid probes are labeled with a detectable molecule. In a second preferred embodiment of said method and kit, said nucleic acid probe or the plurality of nucleic acid probes has been immobilized on a substrate. In a third preferred embodiment, the nucleic acid probe or the plurality of nucleic acid probes comprise either a sequence which is selected from the group consisting of the nucleotide sequences of P1 to P13 and the complementary sequence thereto, B1 to B11, C1 to C11, D1 to D13, E1 to E13 or a biallelic marker selected from the group consisting of A1 to A13 and the complements thereto.

Oligonucleotide arrays

A substrate comprising a plurality of oligonucleotide primers or probes of the invention may be used either for detecting or amplifying targeted sequences in the olfactory receptor gene and may also be used for detecting mutations in the coding or in the non-coding sequences of the olfactory receptor gene.

Any polynucleotide provided herein may be attached in overlapping areas or at random locations on the solid support. Alternatively the polynucleotides of the invention may be attached in an ordered array wherein each polynucleotide is attached to a distinct region of the solid support which does not overlap with the attachment site of any other polynucleotide. Preferably, such an ordered array of polynucleotides is designed to be "addressable" where the distinct locations are recorded and can be accessed as part of an assay procedure. Addressable polynucleotide arrays typically comprise a plurality of different oligonucleotide probes that are coupled to a surface of a substrate in different known locations. The knowledge of the precise location of each polynucleotides location makes these "addressable" arrays particularly us ful in hybridization assays. Any addressable array technology known in the art can be employed with the

polynucleotides of the invention. One particular embodiment of these polynucleotide arrays is known as the Genechips™, and has been generally described in US Patent 5,143,854; PCT publications WO 90/15070 and 92/10092. These arrays may generally be produced using mechanical synthesis methods or light directed synthesis methods which incorporate a combination of photolithographic methods and solid phase oligonucleotide synthesis (Fodor et al., 1991). The immobilization of arrays of oligonucleotides on solid supports has been rendered possible by the development of a technology generally identified as "Very Large Scale Immobilized Polymer Synthesis" (VLSIPS™) in which, typically, probes are immobilized in a high density array on a solid surface of a chip. Examples of VLSIPS™ technologies are provided in US Patents 5,143,854; and 5,412,087 and in PCT Publications WO 90/15070, WO 92/10092 and WO 95/11995, which describe methods for forming oligonucleotide arrays through techniques such as light-directed synthesis techniques. In designing strategies aimed at providing arrays of nucleotides immobilized on solid supports, further presentation strategies were developed to order and display the oligonucleotide arrays on the chips in an attempt to maximize hybridization patterns and sequence information. Examples of such presentation strategies are disclosed in PCT Publications WO 94/12305, WO 94/11530, WO 97/29212 and WO 97/31256.

In another embodiment of the oligonucleotide arrays of the invention, an oligonucleotide probe matrix may advantageously be used to detect mutations occurring in the olfactory receptor gene. For this particular purpose, probes are specifically designed to have a nucleotide sequence allowing their hybridization to the genes that carry known mutations (either by deletion, insertion or substitution of one or several nucleotides). By known mutations, it is meant, mutations on the olfactory receptor gene that have been identified according to, for example, the technique used by Huang et al.(1996) or Samson et al.(1996).

Another technique that is used to detect mutations in the olfactory receptor gene is the use of a high-density DNA array. Each oligonucleotide probe constituting a unit element of the high density DNA array is designed to match a specific subsequence of the olfactory receptor genomic DNA or cDNA. Thus, an array consisting of oligonucleotides complementary to subsequences of the target gene sequence is used to determine the identity of the target sequence with the wild gene sequence, measure its amount, and detect differences between the target sequence and the reference wild gene sequence of the olfactory receptor gene. In one such design, termed 4L tiled array, is implemented a set of four probes (A, C, G, T), preferably 15-nucleotide oligomers. In each set of four probes, the perfect complement will hybridize more strongly than mismatched probes. Consequently, a nucleic acid target of length L is scanned for mutations with a tiled array containing 4L probes, the whole probe set containing all the possible mutations in the known wild reference sequence. The hybridization signals of the 15-mer probe set tiled array are perturbed by a single base change in the target sequence. As a consequence, there is a characteristic loss of signal or a

"footprint" for the probes flanking a mutation position. This technique was described by Chee et al. in 1996.

Consequently, the invention concerns an array of nucleic acid molecules comprising at least one polynucleotide described above as probes and primers. Preferably, the invention concerns an array of nucleic acid comprising at least two polynucleotides described above as probes and primers.

A further object of the invention consists of an array of nucleic acid sequences comprising either at least one of the sequences selected from the group consisting of P1 to P13, B1 to B11, C1 to C11, D1 to D13, E1 to E13, the sequences complementary thereto, a fragment thereof of at least 8, 10, 12, 15, 18, 20, 25, 30, or 40 consecutive nucleotides thereof, and at least one sequence comprising a biallelic marker selected from the group consisting of A1 to A13 and the complements thereto.

The invention also pertains to an array of nucleic acid sequences comprising either at least two of the sequences selected from the group consisting of P1 to P13, B1 to B11, C1 to C11, D1 to D13, E1 to E13, the sequences complementary thereto, a fragment thereof of at least 8 consecutive nucleotides thereof, and at least two sequences comprising a biallelic marker selected from the group consisting of A1 to A13 and the complements thereof.

B. OLF1 TO OFL10 PROTEINS AND POLYPEPTIDE FRAGMENTS

The proteins encoded by the Open Reading Frames of the OLF1 to OLF10 genes are listed individually in the sequence listing as SEQ ID Nos 12-21.

The term "olfactory receptor polypeptides" is used herein to embrace all of the proteins and polypeptides of the present invention. Also forming part of the invention are polypeptides encoded by the polynucleotides of the invention, as well as fusion polypeptides comprising such polypeptides. The invention embodies olfactory receptor proteins from humans, including isolated or purified olfactory receptor proteins consisting of, consisting essentially of, or comprising the sequences of SEQ ID Nos 12-21 or naturally-occurring variants or fragments thereof.

The present invention embodies isolated, purified, and recombinant polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 or 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID Nos 12-21. In a preferred embodiment, the present invention embodies isolated, purified, and recombinant polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 or 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID Nos 12-21 wherein said contiguous span includes at least 1, 2, 3, 5 or 10 of the following amino acid positions in SEQ ID Nos 12-21: 1-20, 21-40, 41-60, 61-80, 81-100, 101-120, 121-140, 141-160, 161-180, 181-200, 201-220, 221-240, 241-260, 261-280, 281-300, 301-the terminal amino acid of the olfactory receptor proteins, to the extent that such amino acid positions are consistent with the lengths of the particular olfactory receptor protein being referred to. In another preferred embodiment, the present invention embodies isolated, purified, and recombinant polypeptides comprising a contiguous span of at least 6

amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a sequence selected from the group consisting of SEQ ID Nos 12, 14, 17, 19 and 21 wherein said contiguous span includes at least 1, 2, 3, 5 or 10 of the following amino acid positions of said selected sequence: 1-10, 11-20, 21-30, 31-40, 41-50, 51-60, 61-70, 71-80, 81-90, 91-100, 101-110, 111-120, 121-130, 131-140, 141-150, 151-160, 161-170, 171-180, 181-190, 191- the terminal amino acid of the olfactory receptor proteins, to the extent that such amino acid positions are consistent with the lengths of the particular olfactory receptor protein being referred to. In further preferred embodiments, the present invention embodies isolated, purified, and recombinant polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 or 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a sequence selected from the group consisting of SEQ ID Nos 12-21, wherein said contiguous span includes at least one amino acid at the following positions of said selected sequence

i) 1-3, 10, 16, 21, 28, 33, 34, 36, 42-44, 46, 49, 53, 54, 57, 59, 63, and 64 for SEQ ID No 12;

ii) 2, 4, 6, 8, 18, 25, 34, 37, 44, 52, 56, 80, 83, 89, 98, 101, 102, 113, 114, 117, 120, 139, 148, 158, 186, 195, 212, 219, 247, 266, 270, 280, 295, 298, 299, 301, 311, and 313-315 for SEQ ID No 13;

iii) 2-4, 6, 18, 21, 25, 34, 37, 98, 99, 102, 113, 114, 133, 143, 148, 158-163, 166, 167, 169, and 170 for SEQ ID No 14;

iv) 2, 4, 6, 8, 18, 25, 34, 37, 44, 52, 54, 56, 80, 83, 89, 98, 101, 102, 113, 114, 117, 120, 139, 148, 158, 186, 195, 212, 219, 247, 266, 270, 280, 298, 299, 311, and 313-315 for SEQ ID No 15;

v) 3, 18, 20, 25, 34, 47, 49, 67, 97, 100, 107, 108, 112, 113, 126, 135, 142, 146, 147, 157, 159-160, 194, 196, 228, 245, 264, 265, 269, 279, 298, and 302 for SEQ ID No 16;

vi) 2, 6, 18, 20, 33, 34, 37, 65, 68, 69, 72, 86, 88, 101, 107, 113, 114, 148, 158, 161, 164, 195, and 198 for SEQ ID No 17;

vii) 2, 6, 7, 52, 56, 67, 88, 94, 97, 110, 113, 116, 119, 120, 127, 135, 150, 153, 164, 174, 175, 180, 184, 217, 221, 259, 261, and 268 for SEQ ID No 18;

viii) 17, 18, 20, 28, 33, 35, 49-52, 105, 111, and 112 for SEQ ID No 19;

ix) 17, 20, 33, 35, 49-53, 56, 111, 112, 132, 138, 141, 147, 154, 157, 160, 163, 164, 194, 197, 204, 211, 214, 218, 219, 252, 265, 286, 295, 301, 303, 305, 306 and 309 for SEQ ID No 20; and

x) 9, 18, 26-28, 34, 47 and 50 for SEQ ID No 21, to the extent that such amino acid lengths are consistent with the lengths of the particular olfactory receptor protein being referred to.

Other preferred OLF1 to OLF10 polypeptide fragments are those located outside the transmembrane domains, most preferably peptide fragments naturally exposed on the cell membrane, particularly those that are available for binding to ligand molecules, either odorant substances or molecules or antibodies directed to the olfactory receptor polypeptides of the invention. Such transmembrane domains TM1 to TM7 are boxed in Figure 1. In other preferred embodiments the contiguous stretch of amino acids comprises the site of a mutation or functional mutation, including a deletion, addition, swap or truncation of the amino acids in the olfactory receptor protein sequence.

The invention also encompasses a purified, isolated, or recombinant polypeptides comprising an amino acid sequence having at least 70, 75, 80, 85, 90, 95, 98 or 99% amino acid identity with the amino acid sequence of SEQ ID Nos 12-21 or a fragment thereof.

The invention also encompasses an olfactory receptor polypeptide or a fragment or a variant thereof in which at least one peptide bound has been modified as defined in the "Definitions" section.

A further object of the invention concerns a purified or isolated polypeptide which is encoded by a nucleic acid comprising a nucleotide sequence selected from the group consisting of SEQ ID Nos 1-11 or fragment or variants thereof.

Such mutated olfactory receptor proteins may be the target of diagnostic tools, such as specific monoclonal or polyclonal antibodies, useful for the detecting the mutated olfactory receptor proteins in a sample.

Olfactory receptor proteins are preferably isolated from human or mammalian tissue samples or expressed from human or mammalian genes.

The olfactory receptor polypeptides of the invention is extracted from cells or tissues of humans or non-human animals. Methods for purifying proteins are known in the art, and include the use of detergents or chaotropic agents to disrupt particles followed by differential extraction and separation of the polypeptides by ion exchange chromatography, affinity chromatography, sedimentation according to density, and gel electrophoresis.

In addition, shorter protein fragments may also be prepared by the conventional methods of chemical synthesis, either in a homogenous solution or in solid phase. As an illustrative embodiment of such chemical polypeptide synthesis techniques, it may be cited the homogenous solution technique described by Houbenweyl in 1974. For solid phase synthesis the technique described by Merrifield (1965) may be used in particular.

Alternatively, the proteins of the invention can be made using routine expression methods known in the art as described below and in the section "Expression of a OLF1 to OLF10 coding polynucleotide ". Briefly, the polynucleotide encoding the desired polypeptide, is ligated into an expression vector suitable for any convenient host. Both eukaryotic and prokaryotic host systems is used in forming recombinant polypeptides. The polypeptide is then isolated from lysed cells or from

the culture medium and purified to the extent needed for its intended use. Purification is by any technique known in the art, for example, differential extraction, salt fractionation, chromatography, centrifugation, and the like. See, for example, Methods in Enzymology for a variety of methods for purifying proteins.

Any olfactory receptor cDNA, including SEQ ID Nos 12-21, may be used to express olfactory receptor proteins and polypeptides. The nucleic acid encoding the olfactory receptor protein or polypeptide to be expressed is operably linked to a promoter in an expression vector using conventional cloning technology. The olfactory receptor insert in the expression vector may comprise the full coding sequence for the olfactory receptor protein or a portion thereof. For example, the olfactory receptor derived insert may encode a polypeptide comprising at least 10 consecutive amino acids of the olfactory receptor protein of SEQ ID Nos 12-21, including any of the polypeptide fragment defined in this section.

The expression vector is any of the mammalian, yeast, insect or bacterial expression systems known in the art. Commercially available vectors and expression systems are available from a variety of suppliers including Genetics Institute (Cambridge, MA), Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence is optimized for the particular expression organism in which the expression vector is introduced, as explained by Hatfield, et al., U.S. Patent No. 5,082,767.

In one embodiment, the entire coding sequence of the olfactory receptor cDNA through the poly A signal of the cDNA are operably linked to a promoter in the expression vector. Alternatively, if the nucleic acid encoding a portion of the olfactory receptor protein lacks a methionine to serve as the initiation site, an initiating methionine can be introduced next to the first codon of the nucleic acid using conventional techniques. Similarly, if the insert from the olfactory receptor cDNA lacks a poly A signal, this sequence can be added to the construct by, for example, splicing out the Poly A signal from pSG5 (Stratagene) using BglI and SalI restriction endonuclease enzymes and incorporating it into the mammalian expression vector pXT1 (Stratagene).

The ligated product is transfected into mouse NIH 3T3 cells using Lipofectin (Life Technologies, Inc., Grand Island, New York) under conditions outlined in the product specification. Positive transfectants are selected after growing the transfected cells in 600ug/ml G418 (Sigma, St. Louis, Missouri).

The above procedures may also be used to express a mutant olfactory receptor protein responsible for a detectable phenotype or a portion thereof.

Purification of the recombinant protein or peptide according to the present invention may be realized by passage onto a Nickel or Copper affinity chromatography column. The Nickel chromatography column may contain the Ni-NTA resin (Porath et al., 1975). The polypeptides or peptides thus obtained may be purified, for example by high performance liquid chromatography,

such as reverse phase and/or cationic exchange HPLC, as described by Rougeot et al. (1994). The reason to prefer this kind of peptide or protein purification is the lack of side products found in the elution samples which renders the resultant purified protein or peptide more suitable for a therapeutic use.

The expressed protein may also be purified using other conventional purification techniques such as ammonium sulfate precipitation or chromatographic separation based on size or charge. The protein encoded by the nucleic acid insert may also be purified using standard immunochromatography techniques. In such procedures, polyclonal or monoclonal antibodies capable of specifically binding to the expressed olfactory receptor protein sof SEQ ID Nos 12-21, or a fragment or a variant thereof, have been previously immobilized onto a chromatography matrix. Such antibodies are described in the section "Antibodies that bind olfactory receptor polypeptides" below. Then, a solution containing the expressed olfactory receptor protein or portion thereof, such as a cell extract, is applied to the chromatography column in conditions allowing the expressed protein to bind to the antibodies in the immunochromatography column. Thereafter, the column is washed to remove non-specifically bound proteins. The specifically bound expressed protein is then released from the column and recovered using standard techniques.

If antibody production is not possible, the nucleic acids encoding the olfactory receptor protein or a portion thereof is incorporated into expression vectors designed for use in purification schemes employing chimeric polypeptides. In such strategies the nucleic acid encoding the olfactory receptor protein or a portion thereof is inserted in frame with the gene encoding the other half of the chimera. The other half of the chimera is β-globin or a nickel binding polypeptide encoding sequence. A chromatography matrix having antibody to β-globin or nickel attached thereto is then used to purify the chimeric protein. Protease cleavage sites is engineered between the β-globin gene or the nickel binding polypeptide and the olfactory receptor protein or portion thereof. Thus, the two polypeptides of the chimera is separated from one another by protease digestion.

One useful expression vector for generating β-globin chimeric proteins is pSG5 (Stratagene), which encodes rabbit β-globin. Intron II of the rabbit β-globin gene facilitates splicing of the expressed transcript, and the polyadenylation signal incorporated into the construct increases the level of expression. These techniques are well known to those skilled in the art of molecular biology. Standard methods are published in methods texts such as Davis et al., (1986) and many of the methods are available from Stratagene, Life Technologies, Inc., or Promega. Polypeptide may additionally be produced from the construct using in vitro translation systems such as the In vitro Express™ Translation Kit (Stratagene).

To confirm expression of the olfactory receptor protein or a portion thereof, the proteins expressed from host cells containing an expression vector containing an insert encoding the olfactory receptor protein or a portion thereof can be compared to the proteins expressed in host cells containing the expression vector without an insert. The presence of a band in samples from cells containing the

expression vector with an insert which is absent in samples from cells containing the expression vector without an insert indicates that the olfactory receptor protein or a portion thereof is being expressed. Generally, the band will have the mobility expected for the olfactory receptor protein or portion thereof. However, the band may have a mobility different than that expected as a result of modifications such as glycosylation, ubiquitination, or enzymatic cleavage.

Other suitable techniques for producing and purifying the olfactory receptor proteins of the invention or their fragments or variants are also described under the heading "Methods for scrreening substances or molecules interacting with an olfactory receptor protein".

Thus, the present invention also concerns a method for the producing a polypeptide of the invention, and especially a polypeptide selected from the group of SEQ ID Nos 12-21 or a fragment or a variant thereof, wherein said methods comprises the steps of :

a) culturing, in an appropriate culture medium, a cell host previously transformed or transfected with the recombinant vector comprising a nucleic acid encoding an olfactory receptor polypeptide of the invention, or a fragment or a variant thereof;

b) harvesting the culture medium thus conditioned or lyze the cell host, for example by sonication or by an osmotic shock;

c) separating or purifying, from the said culture medium, or from the pellet of the resultant host cell lysate the thus produced polypeptide of interest.

d) optionally characterizing the produced polypeptide of interest.

In a specific embodiment of the above method, step a) is preceded by a step wherein the nucleic acid coding for an olfactory receptor polypeptide, or a fragment or a variant thereof, is inserted in an appropriate vector, optionally after an appropriate cleavage of this amplified nucleic acid with one or several restriction endonucleases. The nucleic acid coding for an olfactory receptor polypeptide or a fragment or a variant thereof may be the resulting product of an amplification reaction using a pair of primers according to the invention (by PCR, SDA, TAS, 3SR NASBA, TMA etc.).

C. ANTIBODIES THAT BIND OLFACTORY RECEPTOR POLYPEPTIDES

Any olfactory receptor polypeptide or whole protein may be used to generate antibodies capable of specifically binding to an expressed olfactory receptor protein or fragments thereof as described.

One antibody composition of the invention is capable of specifically binding or specifically bind to the variant of the olfactory receptor protein of SEQ ID Nos 12-21. For an antibody composition to specifically bind to a first variant of olfactory receptor protein, it must demonstrate at least a 5%, 10%, 15%, 20%, 25%, 50%, or 100% greater binding affinity for a first variant of the olfactory receptor protein than for a second variant of the olfactory receptor protein in an ELISA, RIA, or other antibody-based binding assay.

In a preferred embodiment, the invention concerns antibody compositions, either polyclonal or monoclonal, capable of selectively binding, or that selectively bind to an epitope-containing a polypeptide comprising any of the fragments described in the section "OLF1 to OLF10 proteins and polypeptide fragments". Preferred peptide fragments are portions of OLF1 to OLF10 polypeptides that are located outside the transmembrane domains, most preferably peptide fragments naturally exposed on the cell membrane, particularly those that are available for binding to ligand molecules, either odorant substances or molecules or antibodies directed to the olfactory receptor polypeptides of the invention.

The invention also concerns a purified or isolated antibody capable of specifically binding to a mutated olfactory receptor protein or to a fragment or variant thereof comprising an epitope of the mutated olfactory receptor protein. In another preferred embodiment, the present invention concerns an antibody capable of binding to a polypeptide comprising at least 10 consecutive amino acids of an olfactory receptor protein.

In a preferred embodiment, the invention concerns the use in the manufacture of antibodies of a polypeptide comprising any of the fragments described in the section "OLF1 to OLF10 proteins and polypeptide fragments". Preferred peptide fragments are portions of OLF1 to OLF10 polypeptides that are located outside the transmembrane domains, most preferably peptide fragments naturally exposed on the cell membrane, particularly those that are available for recognition of ligand molecules, either odorant substances or molecules or antibodies directed to the olfactory receptor polypeptides of the invention.

The olfactory receptor expressed from a DNA comprising at least one of the nucleic sequences of SEQ ID Nos 1-11 or a fragment or a variant thereof may also be used to generate antibodies capable of specifically binding to the expressed olfactory receptor or fragments or variants thereof. In a preferred embodiment, any of the polynucleotide fragment encoding a polypeptide described in the section " Coding regions of the olfactory receptor gene" may be used to generate such antibodies.

Substantially pure protein or polypeptide is isolated from transfected or transformed cells containing an expression vector encoding the olfactory receptor protein or a portion thereof. The concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms/ml. Monoclonal or polyclonal antibodies to the protein can then be prepared as follows:

1. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes in the olfactory receptor of the present invention or a portion thereof can be prepared from murine hybridomas according to the classical method of Kohler and Milstein, (1975) or derivative methods thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the considered olfactory receptor or a portion thereof over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are

fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, (1980), and derivative methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. et al.

2. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogeneous epitopes in the olfactory receptor of the present invention or a portion thereof can be prepared by immunizing suitable animals with the considered olfactory receptor or a portion thereof, which can be unmodified or modified to enhance immunogenicity. A suitable non-human animal, preferably a non-human mammal, is selected, usually a mouse, rat, rabbit, goat, or horse. Alternatively, a crude preparation which has been enriched for olfactory receptor concentration can be used to generate antibodies. Such proteins, fragments or preparations are introduced into the non-human mammal in the presence of an appropriate adjuvant (e.g. aluminum hydroxide, RIBI, etc.) which is known in the art. In addition the protein, fragment or preparation can be pretreated with an agent which will increase antigenicity, such agents are known in the art and include, for example, methylated bovine serum albumin (mBSA), bovine serum albumin (BSA), Hepatitis B surface antigen, and keyhole limpet hemocyanin (KLH). Serum from the immunized animal is collected, treated and tested according to known procedures. If the serum contains polyclonal antibodies to undesired epitopes, the polyclonal antibodies can be purified by immunoaffinity chromatography.

Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. Techniques for producing and processing polyclonal antisera are known in the art, see for example, Mayer and Walker (1987). An effective immunization protocol for rabbits can be found in Vaitukaitis, J. et al. (1971).

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony et al., (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum. Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, (1980).

Antibody preparations prepared according to either the monoclonal or the polyclonal protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances

in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample. The antibodies may also be used in therapeutic compositions for killing cells expressing the protein or reducing the levels of the protein in the body.

Non-human animals or mammals, whether wild-type or transgenic, which express a different species of olfactory receptor than the one to which antibody binding is desired, and animals which do not express olfactory receptor (i.e. an olfactory receptor knock out animal as described herein) are particularly useful for preparing antibodies. Olfactory receptor knock out animals will recognize all or most of the exposed regions of an olfactory receptor protein as foreign antigens, and therefore produce antibodies with a wider array of olfactory receptor epitopes. Moreover, smaller polypeptides with only 10 to 30 amino acids may be useful in obtaining specific binding to any one of the olfactory receptor proteins. In addition, the humoral immune system of animals which produce a species of olfactory receptor that resembles the antigenic sequence will preferentially recognize the differences between the animal's native olfactory receptor species and the antigen sequence, and produce antibodies to these unique sites in the antigen sequence. Such a technique will be particularly useful in obtaining antibodies that specifically bind to any one of the olfactory receptor proteins.

The present invention also includes, chimeric single-chain Fv antibody fragments (Martineau et al., 1998), antibody fragments obtained through phage display libraries (Ridder et al., 1995; Vaughan et al., 1995) and humanized antibodies (Reinmann et al., 1997; Leger et al., 1997).

The antibodies of the invention may be labeled by any one of the radioactive, fluorescent or enzymatic labels known in the art.

Consequently, the invention is also directed to a method for detecting specifically the presence of a polypeptide according to the invention in a biological sample, said method comprising the following steps :

a) bringing into contact the biological sample with an antibody according to the invention;

b) detecting the antigen-antibody complex formed.

Is also part of the invention a diagnostic kit for in vitro detecting the presence of a polypeptide according to the present invention in a biological sample, wherein said kit comprises:

a) a polyclonal or monoclonal antibody as described above, optionally labeled;

b) a reagent allowing the detection of the antigen-antibody complexes formed, said reagent carrying optionally a label, or being able to be recognized itself by a labeled reagent, more particularly in the case when the above-mentioned monoclonal or polyclonal antibody is not labeled by itself.

D. OLFACTORY RECEPTOR-RELATED BIALLELIC MARKERS

The invention also concerns olfactory receptor-related biallelic markers. As used herein the term "olfactory receptor-related biallelic marker" relates to a set of biallelic markers in linkage

disequilibrium with the olfactory receptor gene. The term olfactory receptor-related biallelic marker includes the biallelic markers designated A1 to A13.

The biallelic markers of the present invention, namely A1 to A13, are disclosed in Table 2 of Example 4. The 13 olfactory receptor-related biallelic markers, A1 to A13, are all located in the genomic non coding regions of the olfactory gene cluster of the invention. Their precise location on the olfactory receptor genomic sequence and their single base polymorphism are indicated in Table 2 and also as features in the sequence listing for SEQ ID No 1. Appropriate pairs of primers allowing the amplification of a nucleic acid containing the polymorphic base of the disclosed olfactory receptor biallelic marker are also listed in Table 1 of Example 3 and in features of SEQ ID No 1.

In the present invention, the biallelic markers can be defined by nucleotide sequences corresponding to oligonucleotides of 47 bases in length comprising at the middle one of the polymorphic base. More particularly, the biallelic markers can be defined by the polynucleotides P1 to P13.

The biallelic markers contained in the olfactory gene cluster of the present invention, or a busset of such biallelic markers, are useful tools to perform association studies, preferably to perform association studies between the statistically significant occurrence of an allele of said biallelic marker in the genome of an individual and a specific phenotype, including a phenotype consisting of an alteration of the olfactory perception of odorant substances or molecules by said individual. The biallelic markers of the invention can also be used, for example, in linkage analysis in which evidence is sought for cosegregation between a locus and a putative trait locus using family studies, such as an alteration of olfactory perception. In addition, the biallellic markers of the invention may be included inthe generation of any complete or partial genetic map of the human genome. These different uses are specifically contemplated in the present invention and claims.

1. Identification of biallelic markers

Any of a variety of methods can be used to screen a genomic fragment for single nucleotide polymorphisms such as differential hybridization with oligonucleotide probes, detection of changes in the mobility measured by gel electrophoresis or direct sequencing of the amplified nucleic acid. A preferred method for identifying biallelic markers involves comparative sequencing of genomic DNA fragments from an appropriate number of unrelated individuals.

In a first embodiment, DNA samples from unrelated individuals are pooled together, following which the genomic DNA of interest is amplified and sequenced. The nucleotide sequences thus obtained are then analyzed to identify significant polymorphisms. One of the major advantages of this method resides in the fact that the pooling of the DNA samples substantially reduces the number of DNA amplification reactions and sequencing reactions, which must be carried out. Moreover, this method is sufficiently sensitive so that a biallelic marker obtained thereby usually shows a sufficient degree of informativeness to be useful in conducting association studies.

In a second embodiment, the DNA samples are not pooled and are therefore amplified and sequenced individually. This method is usually preferred when biallelic markers need to be identified in order to perform association studies within candidate genes. Preferably, highly relevant gene regions such as promoter regions or exon regions may be screened for biallelic markers. A biallelic marker obtained using this method may show a lower degree of informativeness for conducting association studies, e.g. if the frequency of its less frequent allele may be less than about 10%. Such a biallelic marker will, however, be sufficiently informative to conduct association studies and it will further be appreciated that including less informative biallelic markers in the genetic analysis studies of the present invention, may allow in some cases the direct identification of causal mutations, which may, depending on their penetrance, be rare mutations.

The following is a description of the various parameters of a preferred method used by the inventors for the identification of the biallelic markers of the present invention.

Genomic DNA Samples

The genomic DNA samples from which the biallelic markers of the present invention are generated are preferably obtained from unrelated individuals corresponding to a heterogeneous population of known ethnic background. The number of individuals from whom DNA samples are obtained can vary substantially, preferably from about 10 to about 1000, preferably from about 50 to about 200 individuals. It is usually preferred to collect DNA samples from at least about 100 individuals in order to have sufficient polymorphic diversity in a given population to identify as many markers as possible and to generate statistically significant results.

As for the source of the genomic DNA to be subjected to analysis, any test sample can be foreseen without any particular limitation. These test samples include biological samples, which can be tested by the methods of the present invention described herein, and include human and animal body fluids such as whole blood, serum, plasma, cerebrospinal fluid, urine, lymph fluids, and various external secretions of the respiratory, intestinal and genitourinary tracts, tears, saliva, milk, white blood cells, myelomas and the like; biological fluids such as cell culture supernatants; fixed tissue specimens including tumor and non-tumor tissue and lymph node tissues; bone marrow aspirates and fixed cell specimens. The preferred source of genomic DNA used in the present invention is from peripheral venous blood of each donor. Techniques to prepare genomic DNA from biological samples are well known to the skilled technician. Details of a preferred embodiment are provided in Example 2. The person skilled in the art can choose to amplify pooled or unpooled DNA samples.

DNA Amplification

The identification of biallelic markers in a sample of genomic DNA may be facilitated through the use of DNA amplification methods. DNA samples can be pooled or unpooled for the amplification step. DNA amplification techniques are well known to those skilled in the art.

Amplification techniques that can be used in the context of the present invention include, but are not limited to, the ligase chain reaction (LCR) described in EP-A- 320 308, WO 9320227 and EP-A-439 182, the polymerase chain reaction (PCR, RT-PCR) and techniques such as the nucleic acid sequence based amplification (NASBA) described in Guatelli J.C., et al.(1990) and in Compton J.(1991), Q-beta amplification as described in European Patent Application No 4544610, strand displacement amplification as described in Walker et al.(1996) and EP A 684 315 and, target mediated amplification as described in PCT Publication WO 9322461.

LCR and Gap LCR are exponential amplification techniques, both depend on DNA ligase to join adjacent primers annealed to a DNA molecule. In Ligase Chain Reaction (LCR), probe pairs are used which include two primary (first and second) and two secondary (third and fourth) probes, all of which are employed in molar excess to target. The first probe hybridizes to a first segment of the target strand and the second probe hybridizes to a second segment of the target strand, the first and second segments being contiguous so that the primary probes abut one another in 5' phosphate-3'hydroxyl relationship, and so that a ligase can covalently fuse or ligate the two probes into a fused product. In addition, a third (secondary) probe can hybridize to a portion of the first probe and a fourth (secondary) probe can hybridize to a portion of the second probe in a similar abutting fashion. Of course, if the target is initially double stranded, the secondary probes also will hybridize to the target complement in the first instance. Once the ligated strand of primary probes is separated from the target strand, it will hybridize with the third and fourth probes, which can be ligated to form a complementary, secondary ligated product. It is important to realize that the ligated products are functionally equivalent to either the target or its complement. By repeated cycles of hybridization and ligation, amplification of the target sequence is achieved. A method for multiplex LCR has also been described (WO 9320227). Gap LCR (GLCR) is a version of LCR where the probes are not adjacent but are separated by 2 to 3 bases.

For amplification of mRNAs, it is within the scope of the present invention to reverse transcribe mRNA into cDNA followed by polymerase chain reaction (RT-PCR); or, to use a single enzyme for both steps as described in U.S. Patent No. 5,322,770 or, to use Asymmetric Gap LCR (RT-AGLCR) as described by Marshall et al.(1994). AGLCR is a modification of GLCR that allows the amplification of RNA.

The PCR technology is the preferred amplification technique used in the present invention. A variety of PCR techniques are familiar to those skilled in the art. For a review of PCR technology, see White (1997) and the publication entitled "PCR Methods and Applications" (1991, Cold Spring Harbor Laboratory Press). In each of these PCR procedures, PCR primers on either side of the nucleic acid sequences to be amplified are added to a suitably prepared nucleic acid sample along with dNTPs and a thermostable polymerase such as Taq polymerase, Pfu polymerase, or Vent polymerase. The nucleic acid in the sample is denatured and the PCR primers are specifically hybridized to complementary nucleic acid sequences in the sample. The hybridized primers are

extended. Thereafter, another cycle of denaturation, hybridization, and extension is initiated. The cycles are repeated multiple times to produce an amplified fragment containing the nucleic acid sequence between the primer sites. PCR has further been described in several patents including US Patents 4,683,195; 4,683,202; and 4,965,188.

The PCR technology is the preferred amplification technique used to identify new biallelic markers. A typical example of a PCR reaction suitable for the purposes of the present invention is provided in Example 3.

One of the aspects of the present invention is a method for the amplification of the human olfactory receptor gene, particularly of a fragment of the genomic sequence of SEQ ID No 1 or of the coding region sequences of SEQ ID Nos 2-11, or a fragment or a variant thereof in a test sample, preferably using the PCR technology. This method comprises the steps of:

a) contacting a test sample with amplification reaction reagents comprising a pair of amplification primers as described above and located on either side of the polynucleotide region to be amplified, and

b) optionally, detecting the amplification products.

The invention also concerns a kit for the amplification of an olfactory receptor gene sequence, particularly of a portion of the genomic sequence of SEQ ID No 1 or of the coding region sequences of SEQ ID Nos 2-11, or a variant thereof in a test sample, wherein said kit comprises:

a) a pair of oligonucleotide primers located on either side of the olfactory receptor region to be amplified;

b) optionally, the reagents necessary for performing the amplification reaction.

In one embodiment of the above amplification method and kit, the amplification product is detected by hybridization with a labeled probe having a sequence which is complementary to the amplified region. In another embodiment of the above amplification method and kit, primers comprise a sequence which is selected from the group consisting of the nucleotide sequences of B1 to B11, C1 to C11, D1 to D13, and E1 to E13.

In a first embodiment of the present invention, biallelic markers are identified using genomic sequence information generated by the inventors. Sequenced genomic DNA fragments are used to design primers for the amplification of 500 bp fragments. These 500 bp fragments are amplified from genomic DNA and are scanned for biallelic markers. Primers may be designed using the OSP software (Hillier L. and Green P., 1991). All primers may contain, upstream of the specific target bases, a common oligonucleotide tail that serves as a sequencing primer. Those skilled in the art are familiar with primer extensions, which can be used for these purposes.

Sequencing Of Amplified Genomic DNA And Identification Of Single Nucleotide Polymorphisms

The amplification products generated as described above, are then sequenced using any method known and available to the skilled technician. Methods for sequencing DNA using either the dideoxy-mediated method (Sanger method) or the Maxam-Gilbert method are widely known to

those of ordinary skill in the art. Such methods are for example disclosed in Sambrook et al.(1989). Alternative approaches include hybridization to high-density DNA probe arrays as described in Chee et al.(1996).

Preferably, the amplified DNA is subjected to automated dideoxy terminator sequencing reactions using a dye-primer cycle sequencing protocol. Following gel image analysis and DNA sequence extraction, sequence data are automatically processed with adequate software to assess sequence quality.

A polymorphism analysis software is used that detects the presence of biallelic sites among individual or pooled amplified fragment sequences. Polymorphism search is based on the presence of superimposed peaks in the electrophoresis pattern. These peaks which present distinct colors correspond to two different nucleotides at the same position on the sequence. The polymorphism has to be detected on both strands for validation.

Validation Of The Biallelic Markers Of The Present Invention

The polymorphisms are evaluated for their usefulness as genetic markers by validating that both alleles are present in a population. Validation of the biallelic markers is accomplished by genotyping a group of individuals by a method of the invention and demonstrating that both alleles are present. Microsequencing is a preferred method of genotyping alleles. The validation by genotyping step may be performed on individual samples derived from each individual in the group or by genotyping a pooled sample derived from more than one individual. The group can be as small as one individual if that individual is heterozygous for the allele in question. Preferably the group contains at least three individuals, more preferably the group contains five or six individuals, so that a single validation test will be more likely to result in the validation of more of the biallelic markers that are being tested. It should be noted, however, that when the validation test is performed on a small group it may result in a false negative result if as a result of sampling error none of the individuals tested carries one of the two alleles. Thus, the validation process is less useful in demonstrating that a particular initial result is an artifact, than it is at demonstrating that there is a *bona fide* biallelic marker at a particular position in a sequence. All of the genotyping, haplotyping, association, and interaction study methods of the invention may optionally be performed solely with validated biallelic markers.

2. Genotyping of biallelic markers

The polymorphisms identified above can be further confirmed and their respective frequencies can be determined through various methods using the previously described primers and probes. These methods can also be useful for genotyping either new populations in association studies or individuals in the context of detection of alleles of biallelic markers which are known to be associated with a given trait. Those skilled in the art should note that the methods described below can be equally performed on individual or pooled DNA samples.

Once a given polymorphic site has been found and characterized as a biallelic marker as described above, several methods can be used in order to determine the specific allele carried by an individual at the given polymorphic base.

The identification of biallelic markers described previously allows the design of appropriate primers to amplify a region of the olfactory receptor gene cluster containing the polymorphic site of interest and for the detection of such polymorphisms.

Genotyping can be performed using similar methods as those described above for the identification of the biallelic markers, or using other genotyping methods such as those further described below. In preferred embodiments, the comparison of sequences of amplified genomic fragments from different individuals is used to identify new biallelic markers whereas microsequencing is used for genotyping known biallelic markers in diagnostic and genetic analysis applications.

In one embodiment the invention encompasses methods of genotyping comprising determining the identity of a nucleotide at an olfactory receptor-related biallelic marker or the complement thereof in a biological sample; optionally, wherein said olfactory receptor-related biallelic marker is selected from the group consisting of A1 to A13, and the complements thereof; optionally, wherein said biological sample is derived from a single subject; optionally, wherein the identity of the nucleotides at said biallelic marker is determined for both copies of said biallelic marker present in said individual's genome; optionally, wherein said biological sample is derived from multiple subjects; Optionally, the genotyping methods of the invention encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination; Optionally, said method is performed *in vitro*; optionally, further comprising amplifying a portion of said sequence comprising the biallelic marker prior to said determining step; Optionally, wherein said amplifying is performed by PCR, LCR, or replication of a recombinant vector comprising an origin of replication and said fragment in a host cell; optionally, wherein said determining is performed by a hybridization assay, a sequencing assay, a microsequencing assay, or an enzyme-based mismatch detection assay.

Source of Nucleic Acids for genotyping

Any source of nucleic acids, in purified or non-purified form, can be utilized as the starting nucleic acid, provided it contains or is suspected of containing the specific nucleic acid sequence desired. DNA or RNA may be extracted from cells, tissues, body fluids and the like as described above. While nucleic acids for use in the genotyping methods of the invention can be derived from any mammalian source, the test subjects and individuals from which nucleic acid samples are taken are generally understood to be human.

Amplification Of DNA Fragments Comprising Biallelic Markers

Methods and polynucleotides are provided to amplify a segment of nucleotides comprising one or more biallelic marker of the present invention. It will be appreciated that amplification of DNA fragments comprising biallelic markers may be used in various methods and for various purposes and is not restricted to genotyping. Nevertheless, many genotyping methods, although not all, require the previous amplification of the DNA region carrying the biallelic marker of interest. Such methods specifically increase the concentration or total number of sequences that span the biallelic marker or include that site and sequences located either distal or proximal to it. Diagnostic assays may also rely on amplification of DNA segments carrying a biallelic marker of the present invention. Amplification of DNA may be achieved by any method known in the art. Amplification techniques are described above in the section entitled, "DNA amplification."

Some of these amplification methods are particularly suited for the detection of single nucleotide polymorphisms and allow the simultaneous amplification of a target sequence and the identification of the polymorphic nucleotide as it is further described below.

The identification of biallelic markers as described above allows the design of appropriate oligonucleotides, which can be used as primers to amplify DNA fragments comprising the biallelic markers of the present invention. Amplification can be performed using the primers initially used to discover new biallelic markers which are described herein or any set of primers allowing the amplification of a DNA fragment comprising a biallelic marker of the present invention.

In some embodiments the present invention provides primers for amplifying a DNA fragment containing one or more biallelic markers of the present invention. Preferred amplification primers are listed in Example 3. It will be appreciated that the primers listed are merely exemplary and that any other set of primers which produce amplification products containing one or more biallelic markers of the present invention are also of use.

The spacing of the primers determines the length of the segment to be amplified. In the context of the present invention, amplified segments carrying biallelic markers can range in size from at least about 25 bp to 35 kbp. Amplification fragments from 25-3000 bp are typical, fragments from 50-1000 bp are preferred and fragments from 100-600 bp are highly preferred. It will be appreciated that amplification primers for the biallelic markers may be any sequence which allow the specific amplification of any DNA fragment carrying the markers. Amplification primers may be labeled or immobilized on a solid support as described in "Oligonucleotide probes and primers".

Methods of Genotyping DNA samples for Biallelic Markers

Any method known in the art can be used to identify the nucleotide present at a biallelic marker site. Since the biallelic marker allele to be detected has been identified and specified in the present invention, detection will prove simple for one of ordinary skill in the art by employing any of a number of techniques. Many genotyping methods require the previous amplification of the

DNA region carrying the biallelic marker of interest. While the amplification of target or signal is often preferred at present, ultrasensitive detection methods which do not require amplification are also encompassed by the present genotyping methods. Methods well-known to those skilled in the art that can be used to detect biallelic polymorphisms include methods such as, conventional dot blot analyzes, single strand conformational polymorphism analysis (SSCP) described by Orita et al.(1989), denaturing gradient gel electrophoresis (DGGE), heteroduplex analysis, mismatch cleavage detection, and other conventional techniques as described in Sheffield et al.(1991), White et al.(1992), Grompe et al.(1989 and 1993). Another method for determining the identity of the nucleotide present at a particular polymorphic site employs a specialized exonuclease-resistant nucleotide derivative as described in US patent 4,656,127.

Preferred methods involve directly determining the identity of the nucleotide present at a biallelic marker site by sequencing assay, enzyme-based mismatch detection assay, or hybridization assay. The following is a description of some preferred methods. A highly preferred method is the microsequencing technique. The term "sequencing" is generally used herein to refer to polymerase extension of duplex primer/template complexes and includes both traditional sequencing and microsequencing.

1) Sequencing Assays

The nucleotide present at a polymorphic site can be determined by sequencing methods. In a preferred embodiment, DNA samples are subjected to PCR amplification before sequencing as described above. DNA sequencing methods are described in "Sequencing Of Amplified Genomic DNA And Identification Of Single Nucleotide Polymorphisms".

Preferably, the amplified DNA is subjected to automated dideoxy terminator sequencing reactions using a dye-primer cycle sequencing protocol. Sequence analysis allows the identification of the base present at the biallelic marker site.

2) Microsequencing Assays

In microsequencing methods, the nucleotide at a polymorphic site in a target DNA is detected by a single nucleotide primer extension reaction. This method involves appropriate microsequencing primers which, hybridize just upstream of the polymorphic base of interest in the target nucleic acid. A polymerase is used to specifically extend the 3' end of the primer with one single ddNTP (chain terminator) complementary to the nucleotide at the polymorphic site. Next the identity of the incorporated nucleotide is determined in any suitable way.

Typically, microsequencing reactions are carried out using fluorescent ddNTPs and the extended microsequencing primers are analyzed by electrophoresis on ABI 377 sequencing machines to determine the identity of the incorporated nucleotide as described in EP 412 883. Alternatively capillary electrophoresis can be used in order to process a higher number of assays simultaneously. An example of a typical microsequencing procedure that can be used in the context of the present invention is provided in Example 5.

Different approaches can be used for the labeling and detection of ddNTPs. A homogeneous phase detection method based on fluorescence resonance energy transfer has been described by Chen and Kwok (1997) and Chen et al.(1997). In this method, amplified genomic DNA fragments containing polymorphic sites are incubated with a 5'-fluorescein–labeled primer in the presence of allelic dye-labeled dideoxyribonucleoside triphosphates and a modified Taq polymerase. The dye-labeled primer is extended one base by the dye-terminator specific for the allele present on the template. At the end of the genotyping reaction, the fluorescence intensities of the two dyes in the reaction mixture are analyzed directly without separation or purification. All these steps can be performed in the same tube and the fluorescence changes can be monitored in real time. Alternatively, the extended primer may be analyzed by MALDI-TOF Mass Spectrometry. The base at the polymorphic site is identified by the mass added onto the microsequencing primer (see Haff and Smirnov, 1997).

Microsequencing may be achieved by the established microsequencing method or by developments or derivatives thereof. Alternative methods include several solid-phase microsequencing techniques. The basic microsequencing protocol is the same as described previously, except that the method is conducted as a heterogeneous phase assay, in which the primer or the target molecule is immobilized or captured onto a solid support. To simplify the primer separation and the terminal nucleotide addition analysis, oligonucleotides are attached to solid supports or are modified in such ways that permit affinity separation as well as polymerase extension. The 5' ends and internal nucleotides of synthetic oligonucleotides can be modified in a number of different ways to permit different affinity separation approaches, e.g., biotinylation. If a single affinity group is used on the oligonucleotides, the oligonucleotides can be separated from the incorporated terminator regent. This eliminates the need of physical or size separation. More than one oligonucleotide can be separated from the terminator reagent and analyzed simultaneously if more than one affinity group is used. This permits the analysis of several nucleic acid species or more nucleic acid sequence information per extension reaction. The affinity group need not be on the priming oligonucleotide but could alternatively be present on the template. For example, immobilization can be carried out via an interaction between biotinylated DNA and streptavidin-coated microtitration wells or avidin-coated polystyrene particles. In the same manner, oligonucleotides or templates may be attached to a solid support in a high-density format. In such solid phase microsequencing reactions, incorporated ddNTPs can be radiolabeled (Syvänen, 1994) or linked to fluorescein (Livak and Hainer, 1994). The detection of radiolabeled ddNTPs can be achieved through scintillation-based techniques. The detection of fluorescein-linked ddNTPs can be based on the binding of antifluorescein antibody conjugated with alkaline phosphatase, followed by incubation with a chromogenic substrate (such as *p*-nitrophenyl phosphate). Other possible reporter-detection pairs include: ddNTP linked to dinitrophenyl (DNP) and anti-DNP alkaline phosphatase conjugate (Harju et al., 1993) or biotinylated ddNTP and horseradish peroxidase-conjugated

streptavidin with *o*-phenylenediamine as a substrate (WO 92/15712). As yet another alternative solid-phase microsequencing procedure, Nyren et al.(1993) described a method relying on the detection of DNA polymerase activity by an enzymatic luminometric inorganic pyrophosphate detection assay (ELIDA).

Pastinen et al.(1997) describe a method for multiplex detection of single nucleotide polymorphism in which the solid phase minisequencing principle is applied to an oligonucleotide array format. High-density arrays of DNA probes attached to a solid support (DNA chips) are further described below.

In one aspect the present invention provides polynucleotides and methods to genotype one or more biallelic markers of the present invention by performing a microsequencing assay. Preferred microsequencing primers include the nucleotide sequences D1 to Dn and E1 to En. It will be appreciated that the microsequencing primers listed in Example 5 are merely exemplary and that, any primer having a 3' end immediately adjacent to the polymorphic nucleotide may be used. Similarly, it will be appreciated that microsequencing analysis may be performed for any biallelic marker or any combination of biallelic markers of the present invention. One aspect of the present invention is a solid support which includes one or more microsequencing primers listed in Example 5, or fragments comprising at least 8, 12, 15, 20, 25, 30, 40, or 50 consecutive nucleotides thereof, to the extent that such lengths are consistent with the primer described, and having a 3' terminus immediately upstream of the corresponding biallelic marker, for determining the identity of a nucleotide at a biallelic marker site.

3) Mismatch detection assays based on polymerases and ligases

In one aspect the present invention provides polynucleotides and methods to determine the allele of one or more biallelic markers of the present invention in a biological sample, by mismatch detection assays based on polymerases and/or ligases. These assays are based on the specificity of polymerases and ligases. Polymerization reactions places particularly stringent requirements on correct base pairing of the 3' end of the amplification primer and the joining of two oligonucleotides hybridized to a target DNA sequence is quite sensitive to mismatches close to the ligation site, especially at the 3' end. Methods, primers and various parameters to amplify DNA fragments comprising biallelic markers of the present invention are further described above in "Amplification Of DNA Fragments Comprising Biallelic Markers".

Allele Specific Amplification Primers

Discrimination between the two alleles of a biallelic marker can also be achieved by allele specific amplification, a selective strategy, whereby one of the alleles is amplified without amplification of the other allele. For allele specific amplification, at least one member of the pair of primers is sufficiently complementary with a region of an olfactory receptor gene comprising the polymorphic base of a biallelic marker of the present invention to hybridize therewith and to initiate

the amplification. Such primers are able to discriminate between the two alleles of a biallelic marker.

This is accomplished by placing the polymorphic base at the 3' end of one of the amplification primers. Because the extension forms from the 3'end of the primer, a mismatch at or near this position has an inhibitory effect on amplification. Therefore, under appropriate amplification conditions, these primers only direct amplification on their complementary allele. Determining the precise location of the mismatch and the corresponding assay conditions are well within the ordinary skill in the art.

Ligation/Amplification Based Methods

The "Oligonucleotide Ligation Assay" (OLA) uses two oligonucleotides which are designed to be capable of hybridizing to abutting sequences of a single strand of a target molecules. One of the oligonucleotides is biotinylated, and the other is detectably labeled. If the precise complementary sequence is found in a target molecule, the oligonucleotides will hybridize such that their termini abut, and create a ligation substrate that can be captured and detected. OLA is capable of detecting single nucleotide polymorphisms and may be advantageously combined with PCR as described by Nickerson et al.(1990). In this method, PCR is used to achieve the exponential amplification of target DNA, which is then detected using OLA.

Other amplification methods which are particularly suited for the detection of single nucleotide polymorphism include LCR (ligase chain reaction), Gap LCR (GLCR) which are described above in "DNA Amplification". LCR uses two pairs of probes to exponentially amplify a specific target. The sequences of each pair of oligonucleotides, is selected to permit the pair to hybridize to abutting sequences of the same strand of the target. Such hybridization forms a substrate for a template-dependant ligase. In accordance with the present invention, LCR can be performed with oligonucleotides having the proximal and distal sequences of the same strand of a biallelic marker site. In one embodiment, either oligonucleotide will be designed to include the biallelic marker site. In such an embodiment, the reaction conditions are selected such that the oligonucleotides can be ligated together only if the target molecule either contains or lacks the specific nucleotide that is complementary to the biallelic marker on the oligonucleotide. In an alternative embodiment, the oligonucleotides will not include the biallelic marker, such that when they hybridize to the target molecule, a "gap" is created as described in WO 90/01069. This gap is then "filled" with complementary dNTPs (as mediated by DNA polymerase), or by an additional pair of oligonucleotides. Thus at the end of each cycle, each single strand has a complement capable of serving as a target during the next cycle and exponential allele-specific amplification of the desired sequence is obtained.

Ligase/Polymerase-mediated Genetic Bit Analysis™ is another method for determining the identity of a nucleotide at a preselected site in a nucleic acid molecule (WO 95/21271). This method involves the incorporation of a nucleoside triphosphate that is complementary to the nucleotide

present at the preselected site onto the terminus of a primer molecule, and their subsequent ligation to a second oligonucleotide. The reaction is monitored by detecting a specific label attached to the reaction's solid phase or by detection in solution.

4) Hybridization Assay Meth ds

A preferred method of determining the identity of the nucleotide present at a biallelic marker site involves nucleic acid hybridization. The hybridization probes, which can be conveniently used in such reactions, preferably include the probes defined herein. Any hybridization assay may be used including Southern hybridization, Northern hybridization, dot blot hybridization and solid-phase hybridization (see Sambrook et al., 1989).

Hybridization refers to the formation of a duplex structure by two single stranded nucleic acids due to complementary base pairing. Hybridization can occur between exactly complementary nucleic acid strands or between nucleic acid strands that contain minor regions of mismatch. Specific probes can be designed that hybridize to one form of a biallelic marker and not to the other and therefore are able to discriminate between different allelic forms. Allele-specific probes are often used in pairs, one member of a pair showing perfect match to a target sequence containing the original allele and the other showing a perfect match to the target sequence containing the alternative allele. Hybridization conditions should be sufficiently stringent that there is a significant difference in hybridization intensity between alleles, and preferably an essentially binary response, whereby a probe hybridizes to only one of the alleles. Stringent, sequence specific hybridization conditions, under which a probe will hybridize only to the exactly complementary target sequence are well known in the art (Sambrook et al., 1989). Stringent conditions are sequence dependent and will be different in different circumstances. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. Although such hybridization can be performed in solution, it is preferred to employ a solid-phase hybridization assay. The target DNA comprising a biallelic marker of the present invention may be amplified prior to the hybridization reaction. The presence of a specific allele in the sample is determined by detecting the presence or the absence of stable hybrid duplexes formed between the probe and the target DNA. The detection of hybrid duplexes can be carried out by a number of methods. Various detection assay formats are well known which utilize detectable labels bound to either the target or the probe to enable detection of the hybrid duplexes. Typically, hybridization duplexes are separated from unhybridized nucleic acids and the labels bound to the duplexes are then detected. Those skilled in the art will recognize that wash steps may be employed to wash away excess target DNA or probe as well as unbound conjugate. Further, standard heterogeneous assay formats are suitable for detecting the hybrids using the labels present on the primers and probes.

Two recently developed assays allow hybridization-based allele discrimination with no need for separations or washes (see Landegren U. et al., 1998). The TaqMan assay takes advantage of the 5' nuclease activity of Taq DNA polymerase to digest a DNA probe annealed specifically to the

accumulating amplification product. TaqMan probes are labeled with a donor-acceptor dye pair that interacts via fluorescence energy transfer. Cleavage of the TaqMan probe by the advancing polymerase during amplification diss ciates the donor dye from the quenching acceptor dye, greatly increasing the donor fluorescence. All reagents necessary to detect two allelic variants can be assembled at the beginning of the reaction and the results are monitored in real time (see Livak et al., 1995). In an alternative homogeneous hybridization based procedure, molecular beacons are used for allele discriminations. Molecular beacons are hairpin-shaped oligonucleotide probes that report the presence of specific nucleic acids in homogeneous solutions. When they bind to their targets they undergo a conformational reorganization that restores the fluorescence of an internally quenched fluorophore (Tyagi et al., 1998).

The polynucleotides provided herein can be used to produce probes which can be used in hybridization assays for the detection of biallelic marker alleles in biological samples. These probes are characterized in that they preferably comprise between 8 and 50 nucleotides, and in that they are sufficiently complementary to a sequence comprising a biallelic marker of the present invention to hybridize thereto and preferably sufficiently specific to be able to discriminate the targeted sequence for only one nucleotide variation. A particularly preferred probe is 25 nucleotides in length. Preferably the biallelic marker is within 4 nucleotides of the center of the polynucleotide probe. In particularly preferred probes, the biallelic marker is at the center of said polynucleotide. Preferred probes comprise a nucleotide sequence selected from the group consisting of amplicons listed in Table 1 and the sequences complementary thereto, or a fragment thereof, said fragment comprising at least about 8 consecutive nucleotides, preferably 10, 15, 20, more preferably 25, 30, 40, 47, or 50 consecutive nucleotides and containing a polymorphic base. Preferred probes comprise a nucleotide sequence selected from the group consisting of P1 to P13 and the sequences complementary thereto. In preferred embodiments the polymorphic base(s) are within 5, 4, 3, 2, 1, nucleotides of the center of the said polynucleotide, more preferably at the center of said polynucleotide.

Preferably the probes of the present invention are labeled or immobilized on a solid support. Labels and solid supports are further described in "Oligonucleotide Probes and Primers". The probes can be non-extendable as described in "Oligonucleotide Probes and Primers".

By assaying the hybridization to an allele specific probe, one can detect the presence or absence of a biallelic marker allele in a given sample. High-Throughput parallel hybridization in array format is specifically encompassed within "hybridization assays" and are described below.

5) Hybridization To Addressable Arrays Of Oligonucleotides

Hybridization assays based on oligonucleotide arrays rely on the differences in hybridization stability of short oligonucleotides to perfectly matched and mismatched target sequence variants. Efficient access to polymorphism information is obtained through a basic structure comprising high-density arrays of oligonucleotide probes attached to a solid support (e.g., the chip) at selected

positions. Each DNA chip can contain thousands to millions of individual synthetic DNA probes arranged in a grid-like pattern and miniaturized to the size of a dime.

The chip technology has already been applied with success in numerous cases. For example, the screening of mutations has been undertaken in the BRCA1 gene, in *S. cerevisiae* mutant strains, and in the protease gene of HIV-1 virus (Hacia et al., 1996; Shoemaker et al., 1996; Kozal et al., 1996). Chips of various formats for use in detecting biallelic polymorphisms can be produced on a customized basis by Affymetrix (GeneChip™), Hyseq (HyChip and HyGnostics), and Protogene Laboratories.

In general, these methods employ arrays of oligonucleotide probes that are complementary to target nucleic acid sequence segments from an individual which, target sequences include a polymorphic marker. EP 785280 describes a tiling strategy for the detection of single nucleotide polymorphisms. Briefly, arrays may generally be "tiled" for a large number of specific polymorphisms. By "tiling" is generally meant the synthesis of a defined set of oligonucleotide probes which is made up of a sequence complementary to the target sequence of interest, as well as preselected variations of that sequence, e.g., substitution of one or more given positions with one or more members of the basis set of nucleotides. Tiling strategies are further described in PCT application No. WO 95/11995. In a particular aspect, arrays are tiled for a number of specific, identified biallelic marker sequences. In particular, the array is tiled to include a number of detection blocks, each detection block being specific for a specific biallelic marker or a set of biallelic markers. For example, a detection block may be tiled to include a number of probes, which span the sequence segment that includes a specific polymorphism. To ensure probes that are complementary to each allele, the probes are synthesized in pairs differing at the biallelic marker. In addition to the probes differing at the polymorphic base, monosubstituted probes are also generally tiled within the detection block. These monosubstituted probes have bases at and up to a certain number of bases in either direction from the polymorphism, substituted with the remaining nucleotides (selected from A, T, G, C and U). Typically the probes in a tiled detection block will include substitutions of the sequence positions up to and including those that are 5 bases away from the biallelic marker. The monosubstituted probes provide internal controls for the tiled array, to distinguish actual hybridization from artefactual cross-hybridization. Upon completion of hybridization with the target sequence and washing of the array, the array is scanned to determine the position on the array to which the target sequence hybridizes. The hybridization data from the scanned array is then analyzed to identify which allele or alleles of the biallelic marker are present in the sample. Hybridization and scanning may be carried out as described in PCT application No. WO 92/10092 and WO 95/11995 and US patent No. 5,424,186.

Thus, in some embodiments, the chips may comprise an array f nucleic acid sequences of fragments of about 15 nucleotides in length. In further embodiments, the chip may comprise an array including at least one of the sequences selected from the group consisting of amplicons listed

in table 1 and the sequences complementary thereto, or a fragment thereof, said fragment comprising at least about 8 consecutive nucleotides, preferably 10, 15, 20, more preferably 25, 30, 40, 47, or 50 consecutive nucleotides and containing a polymorphic base. In preferred embodiments the polymorphic base is within 5, 4, 3, 2, 1, nucleotides of the center of the said polynucleotide, more preferably at the center of said polynucleotide. In some embodiments, the chip may comprise an array of at least 2, 3, 4, 5, 6, 7, 8 or more of these polynucleotides of the invention. Solid supports and polynucleotides of the present invention attached to solid supports are further described in "Oligonucleotide Probes And Primers".

6) Integrated Systems

Another technique, which may be used to analyze polymorphisms, includes multicomponent integrated systems, which miniaturize and compartmentalize processes such as PCR and capillary electrophoresis reactions in a single functional device. An example of such technique is disclosed in US patent 5,589,136 which describes the integration of PCR amplification and capillary electrophoresis in chips.

Integrated systems can be envisaged mainly when microfluidic systems are used. These systems comprise a pattern of microchannels designed onto a glass, silicon, quartz, or plastic wafer included on a microchip. The movements of the samples are controlled by electric, electroosmotic or hydrostatic forces applied across different areas of the microchip to create functional microscopic valves and pumps with no moving parts.

For genotyping biallelic markers, the microfluidic system may integrate nucleic acid amplification, microsequencing, capillary electrophoresis and a detection method such as laser-induced fluorescence detection.

E. EXPRESSION OF AN OL1 TO OLF10 CODING POLYNUCLEOTIDE

Any of the coding polynucleotides of the invention may be inserted into recombinant vectors for expression in a recombinant host cell or a recombinant host organism.

Thus, the present invention also encompasses a family of recombinant vectors that contains a coding polynucleotide from the group of coding polynucleotides OLF1 to OLF10 genes. Consequently, the present invention further deals with a recombinant vector comprising a polynucleotide comprising any of the coding sequence of SEQ ID No 1, preferably those selected from the group consisting of SEQ ID Nos 2-11.

In a first preferred embodiment, the present invention relates to expression vectors which include nucleic acids encoding an olfactory receptor protein described herein under the control of an exogenous regulatory sequence.

In a second preferred embodiment, a recombinant vector of the invention is used to amplify the inserted p lynucleotide derived from an olfactory receptor genomic sequence selected from the group consisting of the nucleic acids of SEQ ID No 1 and of olfactory receptor cDNAs, for example

the open reading frames of SEQ ID Nos 2-11, in a suitable cell host , this polynucleotide being amplified at every time that the recombinant vector replicates.

More particularly, the present invention relates to expression vectors which include nucleic acids encoding an olfactory receptor protein, preferably the olfactory receptor proteins of the amino acid sequence of SEQ ID Nos 12-21 or variants or fragments thereof, under the control of an exogenous regulatory sequence.

Generally, a recombinant vector of the invention may comprise any of the polynucleotides described herein, including regulatory sequences, and coding sequences, as well as any olfactory receptor primer or probe as defined above. More particularly, the recombinant vectors of the present invention can comprise any of the polynucleotides described in the "Coding Regions of the olfactory receptor gene" section, "Genomic sequence of the olfactory receptor gene" section, the "Oligonucleotide Probes And Primers" section and the "Polynucleotide constructs" section.

Some of the elements which can be found in the vectors of the present invention are described in further detail in the following sections.

Vectors

A recombinant vector according to the invention comprises, but is not limited to, a YAC (Yeast Artificial Chromosome), a BAC (Bacterial Artificial Chromosome), a phage, a phagemid, a cosmid, a plasmid or even a linear DNA molecule which may consist of a chromosomal, non-chromosomal and synthetic DNA. Such a recombinant vector can comprise a transcriptional unit comprising an assembly of

(1) a genetic element or elements having a regulatory role in gene expression, for example promoters or enhancers. Enhancers are cis-acting elements of DNA, usually from about 10 to 300 bp that act on the promoter to increase the transcription.

(2) a structural or coding sequence which is transcribed into mRNA and eventually translated into a polypeptide, and

(3) appropriate transcription initiation and termination sequences. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an N-terminal residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

Generally, recombinant expression vectors will include origins of replication, selectable markers permitting transformation of the host cell, and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium.

The selectable marker genes for selection of transformed host cells are preferably dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, TRP1 for *S. cerevisiae* or tetracycline, rifampicine or ampicillin resistance in *E. coli*, or levan saccharase for mycobacteria.

For facilitating the purification of the expressed protein and increasing its stability, the coding sequence of an olfactory receptor according to the invention can be fused in its N- or C-terminus with protein such as MBP (maltose binding protein) and GST (Glutathione S transferase) or with tag such as poly-histidine tag, Strep tag, Bio tag, and flag peptide epitope tag, those being detailed below. Thioredoxin can be eventually inserted between the olfactory receptor and the tag.

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired polypeptide with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia, Uppsala, Sweden), and GEM1 (Promega Biotec, Madison, WI, USA).

Large numbers of suitable vectors and promoters are known to those of skill in the art, and commercially available, such as bacterial vectors : pQE70, pQE60, pQE-9 (Qiagen), pbs, pD10, phagescript, psiX174, pbluescript SK, pbsks, pNH8A, pNH16A, pNH18A, pNH46A (Stratagene); ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia); or eukaryotic vectors : pWLNEO, pSV2CAT, pOG44, pXT1, pSG (Stratagene); pSVK3, pBPV, pMSG, pSVL (Pharmacia); baculovirus transfer vector pVL1392/1393 (Pharmingen); pQE-30 (QIAexpress).

A suitable vector for the expression of the olfactory receptor above-defined or their peptide fragments is baculovirus vector that can be propagated in insect cells and in insect cell lines. A specific suitable host vector system is the pVL1392/1393 baculovirus transfer vector (Pharmingen) that is used to transfect the SF9 cell line (ATCC N°CRL 1711) which is derived from *Spodoptera frugiperda*.

Other suitable vectors for the expression of an olfactory receptor or their peptide fragments or variants in a baculovirus expression system include those described by Chai et al. (1993), Vlasak et al. (1983) and Lenhard et al. (1996).

Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example SV40 origin, early promoter,

enhancer, splice and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

Promoters

The suitable promoter regions used in the expression vectors according to the present invention are chosen taking into account of the cell host in which the heterologous gene has to be expressed.

A suitable promoter may be heterologous with respect to the nucleic acid for which it controls the expression or alternatively can be endogenous to the native polynucleotide containing the coding sequence to be expressed. Additionally, the promoter is generally heterologous with respect to the recombinant vector sequences within which the construct promoter/coding sequence has been inserted.

Thus, the promoter is selected among the group comprising :

- an internal or an endogenous promoter, such as the natural promoter associated with the structural gene coding for the desired olfactory receptor polypeptide or the fragment or variant thereof; such a promoter may be completed by a regulatory element derived from the vertebrate host, in particular an activator element;

- a promoter derived from a cytoskeletal protein gene such as the desmin promoter (Bolmont et al., 1990; Zhenlin et al., 1989) or a promoter derived from a gene specifically expressed in epithelial cells and most preferably in olfactory epithelial cells.

Preferred bacterial promoters are the LacI, LacZ, the T3 or T7 bacteriophage RNA polymerase promoters, the polyhedrin promoter, or the p10 protein promoter from baculovirus (Kit Novagen) (Smith et al., 1983.; O'Reilly et al., 1992), the lambda P_R promoter or also the trc promoter.

Promoter regions can be selected from any desired gene using, for example, CAT (chloramphenicol transferase) vectors and more preferably pKK232-8 and pCM7 vectors. Particularly preferred bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, PL and trp. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-L. Selection of a convenient vector and promoter is well within the level of ordinary skill in the art.

The choice of a determined promoter, among the above-described promoters is well in the ability of one skill in the art, guided by his knowledge in the genetic engineering technical field, and by being also guided by the book of Sambrook et al. in 1989 or also by the procedures described by Fuller et al. in 1996 (Fuller S.A. et al., 1996).

A preferred constitutive promoter that is used is one of the internal promoters that are active in the resting fibroblasts such the promoter of the phosphoglycerate kinase gene (PGK-1). The PGK-1 promoter is either the mouse promoter or the human promoter such as described by Adra et al.(

1987). Other constitutive promoters may also be used such that the beta-actin promoter (Kort et al., 1983) or the vimentin promoter (Rettlez and Basenga, 1987).

The vector containing the appropriate DNA sequence as described above, more preferably a OLF1 to OLF10 coding polynucleotide, can be utilized to transform an appropriate host to allow the expression of the desired polypeptide or polynucleotide.

Other types of vectors

The *in vivo* expression of an olfactory receptor polypeptide encompassed by the invention or a fragment or a variant thereof may be useful in order to correct a genetic defect related to the expression of the native gene in a host organism or to the production of biologically active olfactory receptor proteins.

Consequently, the present invention also deals with recombinant expression vectors mainly designed for the *in vivo* production of a therapeutic peptide fragment by the introduction of the genetic information in the organism of the patient to be treated. This genetic information may be introduced *in vitro* in a cell that has been previously extracted from the organism, the modified cell being subsequently reintroduced in the said organism, directly *in vivo* into the appropriate tissue, and preferably in the olfactory epithelium.

One specific embodiment for a method for delivering the corresponding protein or peptide to the interior of a cell of a vertebrate *in vivo* comprises the step of introducing a preparation comprising a physiologically acceptable carrier and a naked polynucleotide operatively coding for the polypeptide into the interstitial space of a tissue comprising the cell, whereby the naked polynucleotide is taken up into the interior of the cell and has a physiological effect.

In a specific embodiment, the invention provides a composition for the *in vivo* production of an olfactory receptor polypeptide described therein containing a naked polynucleotide operatively coding for an olfactory receptor selected from the group of OLF1 to OLF10 or a fragment or a variant thereof, in solution in a physiologically acceptable carrier and suitable for introduction into a tissue to cause cells of the tissue to express the said protein or polypeptide.

Advantageously, the composition described above is administered locally, near the site in which the expression of the olfactory receptor polypeptide under consideration or a fragment or a variant thereof is sought.

The polynucleotide operatively coding for an olfactory receptor polypeptide or a fragment or variant thereof may be a vector comprising the genomic DNA or the complementary DNA (cDNA) coding for the corresponding protein and a promoter sequence allowing the expression of the genomic DNA or the complementary DNA in the desired eukaryotic cells, such as vertebrate cells, specifically mammalian cells.

This vector may also contain one origin of replication that allows it to replicate in the eukaryotic host cell such as an origin of replication from a bovine papillomavirus. Alternatively, the vector can contain several, for example two, origins of replication of different origins in order to

allow said vector to replicate in different host cells, typically both in a prokaryotic cell such as *E. coli* and in an eukaryotic cell such as a mammalian epithelial cell, preferably a mammalian olfactory epithelial cell.

Compositions comprising a polynucleotide are described in the PCT application N° WO 90/11092 (Vical Inc.) and also in the PCT application N° WO 95/11307 (Institut Pasteur, INSERM, Université d'Ottawa) as well as in the articles of Tacson et al. (1996) and of Huygen et al. (1996).

In another embodiment, the DNA to be introduced is complexed with DEAE-dextran (Pagano et al., 1967) or with nuclear proteins (Kaneda et al., 1989), with lipids (Felgner et al., 1987) or encapsulated within liposomes (Fraley et al., 1980).

In another embodiment, the polynucleotide encoding an olfactory receptor polypeptide of the invention or a fragment or a variant thereof may be included in a transfection system comprising polypeptides that promote its penetration within the host cells as it is described in the PCT application WO 95/10534 (Seikagaku Corporation). They can also be encapsulated in polymer microparticles as it is described in the PCT Application No WO 94/27238.

The vector according to the present invention may advantageously be administered in the form of a gel that facilitates their transfection into the cells. Such a gel composition may be a complex of poly-L-lysine and lactose, as described by Midoux (1993) or also poloxamer 407 as described by Pastore (1994). Said vector may also be suspended in a buffer solution or be associated with liposomes.

The amount of the vector to be injected to the desired host organism vary according to the site of injection. As an indicative dose, it will be injected between 0,1 and 100 µg of the vector in an animal body, preferably a mammal body, for example a mouse body.

In another embodiment of the vector according to the invention, said vector may be introduced in vitro in a host cell, preferably in a host cell previously harvested from the animal to be treated and more preferably a somatic cell such as a muscle cell. In a subsequent step, the cell that has been transformed with the vector coding for the desired olfactory receptor polypeptide or the desired fragment or variant thereof is implanted back into the animal body in order to deliver the recombinant protein within the body either locally or systemically.

Suitable vectors for the *in vivo* expression of an olfactory receptor polypeptide of the invention or a fragment or a variant thereof are described hereunder.

In one specific embodiment, the vector is derived from an adenovirus. Preferred adenoviruses vectors according to the invention are those described by Feldman and Steg (1996) or Ohno et al. (1994). Another preferred recombinant adenovirus according to this specific embodiment of the present invention is the adenovirus described by Ohwada et al. (1996) or the human adenovirus type 2 or 5 (Ad 2 or Ad 5) or an adenovirus of animal origin (French patent application N° FR-93.05954).

Among the adenoviruses of animal origin it can be cited the adenoviruses of canine (CAV2, strain Manhattan or A26/61[ATCC VR-800]), bovine, murine (Mav1, Beard et al., 1980) or simian (SAV).

Preferably, the inventors are using recombinant defective adenoviruses that may be prepared following a technique well-known by one skill in the art, for example as described by Levrero et al. (1991) or by Graham (1984) or in the European patent application N° EP-185.573. Another defective recombinant adenovirus that may be used according to the present invention, as well as a composition of matter containing such a defective recombinant adenovirus, is described in the PCT application N° WO 95/14785.

Retrovirus vectors and adeno-associated virus vectors are generally understood to be the recombinant gene delivery system of choice for the transfer of exogenous polynucleotides *in vivo* , particularly to mammals, including humans. These vectors provide efficient delivery of genes into cells, and the transferred nucleic acids are stably integrated into the chromosomal DNA of the host.

The use of recombinant retrovirus vectors containing a nucleic acid according to the invention is also encompassed within the scope of the invention. A major prerequisite for the use of retroviruses is to ensure the safety of their use, particularly with regard to the possibility of the spread of wild-type virus in the cell population. The development of specialized cell lines (termed "packaging cells") which produce only replication defective retroviruses has increased the utility of retroviruses for *in vivo* gene delivery, and defective retroviruses are well characterized for use in gene transfer. Thus, recombinant retroviruses can be constructed in which a part of the retroviral coding sequence (*gag*, *pol*, *env*) has been replaced by nucleic acid encoding an olfactory receptor rendering the retrovirus defective. Protocols for producing recombinant retroviruses and for infecting cells *in vitro* and *in vivo* with such viruses can be found in "Current Protocols in Molecular Biology" (1989).

Furthermore, it has been shown that it is possible to limit the infection spectrum of retroviruses and consequently of retroviral-based vectors, by modifying the viral packaging proteins on the surface of the viral particle, as described for example in the PCT Application No WO 93/25234 or in the PCT Application No WO 94/ 06920. For instance, strategies for the modification of the infection spectrum of retroviral vectors include : coupling antibodies specific for cell surface antigens to the viral *env* protein (Julan et al., 1992) or coupling cell surface receptor ligands to the viral *env* protein (Neda et al., 1991). Coupling can be in the form of the chemical cross-linking with a protein or other variety (e.g. lactose to convert the *env* protein to an asialoglycoprotein), as well by generating fusion proteins (e.g. single-chain antibody/*env* fusion proteins). This technique, while useful to limit or otherwise direct the infection to certain tissue types, can also be used to convert an ecotropic vector into an amphotropic vector.

Particularly preferred retroviruses for the preparation or construction of retroviral *in vitro* or *in vitro* gene delivery vehicles of the present invention include retroviruses selected from the group

consisting of Mink-Cell Focus Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis virus and Rous Sarcoma virus. Particularly preferred Murine Leukemia Viruses include 4070A and 1504A (Hartley et al., 1976), Abelson (ATCC No VR-999), Friend (ATCC No VR-245), Gross (ATCC No VR-590), Rauscher (ATCC No VR-998) and Moloney Murine Leukemia Virus (ATCC No VR-190; PCT Application No WO 94/24298). Particularly preferred Rous Sarcoma Viruses include Bryan high titer (ATCC Nos VR-334, VR-657, VR-726, VR-659 and VR-728). Another preferred retroviral vector is that described by Roth et al. (Roth J.A. et al., 1996).

Yet another viral vector system that is contemplated by the invention consists in the adeno-associated virus (AAV). Adeno-associated virus is a naturally occurring defective virus that requires another virus, such as an adenovirus or a herpes virus, as a helper virus for efficient replication and a productive life cycle (Muzyczka et al., 1992). It is also one of the few viruses that may integrate its DNA into non-dividing cells, and exhibits a high frequency of stable integration (Flotte et al., 1992; Samulski et al., 1989; McLaughlin et al., 1989). One advantageous feature of AAV derives from its reduced efficacy for transducing primary cells relative to transformed cells.

Cell hosts

Another object of the invention consists in host cell that have been transformed or transfected with one of the polynucleotides described therein, and more precisely a polynucleotide comprising the coding sequence of any of the olfactory receptor polypeptide having the amino acid sequence of SEQ ID Nos 12-21 or fragments or variants thereof. Are included host cells that are transformed (prokaryotic cells) or that are transfected (eukaryotic cells) with a recombinant vector such as one of those described above.

A recombinant host cell of the invention comprises any one of the polynucleotides or the recombinant vectors described therein. More particularly, the cell hosts of the present invention can comprise any of the polynucleotides described in the "Coding regions of the olfactory receptor gene" section, "Genomic sequence of olfactory receptor gene " section, the "Oligonucleotide Probes And Primers" section, the "Polynucleotide constructs" section.and the " Expression of an OLF1 to OLF10 coding polypeptide" section.

Suitable prokaryotic hosts for transformation include *E. coli*, *Bacillus subtilis*, as well as various species within the genera of *Streptomyces* or *Mycobacterium*. Suitable eukaryotic hosts comprise yeast, insect cells, such as Drosophila and Sf9. Various mammalian cell hosts can also be employed to express recombinant protein. Examples of mammalian cell hosts include the COS-7 lines of monkey kidney fibroblasts (Guzman, 1981), and other cell lines capable of expressing a compatible vector, for example the C127, 3T3, CHO, HeLa and BHK cell lines. The selection of an host is within the scope of the one skilled in the art.

Preferred cell hosts used as recipients for the expression vectors of the invention are the followings :

a) Prokaryotic host cells : *Escherichia coli* strains (I.E. DH5-α strain) or *Bacillus subtilis*.

b) Eukaryotic host cells : HeLa cells (ATCC N°CCL2; N°CCL2.1; N°CCL2.2). Cv 1 cells (ATCC N°CCL70), COS cells (ATCC N°CRL1650; N°CRL1651), Sf-9 cells (ATCC N°CRL1711).

The constructs in the host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence.

Following transformation of a suitable host and growth of the host to an appropriate cell density, the selected promoter is induced by appropriate means, such as temperature shift or chemical induction, and cells are cultivated for an additional period.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Such methods are well known by the skill artisan.

Transgenic animals

The terms "transgenic animals" or "host animals" are used herein designate animals that have their genome genetically and artificially manipulated so as to include one of the nucleic acids according to the invention. Preferred animals are non-human mammals and include those belonging to a genus selected from *Mus* (e.g. mice), *Rattus* (e.g. rats) and *Oryctogalus* (e.g. rabbits) which have their genome artificially and genetically altered by the insertion of a nucleic acid according to the invention.

The transgenic animals of the invention all include within a plurality of their cells a cloned recombinant or synthetic DNA sequence, more specifically one of the purified or isolated nucleic acids comprising an olfactory receptor coding sequence selected from the group OLF1 to OLF10 an olfactory receptor regulatory polynucleotide or a DNA sequence encoding an antisense polynucleotide such as described in the present specification.

More particularly, transgenic animals according to the invention contain in their somatic cells and/or in their germ line cells any of the polynucleotides described in the "Coding regions of the olfactory receptor gene" section, "Genomic sequence of olfactory receptor gene " section, the "Oligonucleotide Probes And Primers" section, the "Polynucleotide constructs" section and the " Expression of an OLF1 to OLF10 coding polypeptide" section.

The replacement of the native genomic olfactory receptor sequence by a defective copy of said sequence may be preformed by techniques of gene targeting. Such techniques are notably described by Burright et al. (1997), Bates et al. (1997), Mangiarini et al. (1997), Davies et al. (1997).

Second preferred transgenic animals of the invention have the murine olfactory receptor gene replaced either by a defective copy of the murine olfactory receptor gene or by an interrupted copy of the human olfactory receptor gene. A "defective copy" of a murine or a human olfactory receptor gene, is intended to designate a modified copy of these genes that is not or poorly transcribed in the resulting recombinant host animal or a modified copy of these genes leading to the

absence of synthesis of the corresponding translation product or alternatively leading to a modified and/or truncated translation product lacking the biological activity of the wild type olfactory receptor protein. The altered translation product thus contains amino acid modifications, deletions and substitutions. Modifications and deletions may render the naturally occurring gene nonfunctional, thus leading to a "knockout animal". These transgenic animals are critical for the creation of animal models of human diseases, and for eventual treatment of disorders related to alteration of the olfactory perception of odorant substances or molecules. Examples of such knockout mice are described in the PCT Applications Nos WO 97/34641, WO 96/12792 and WO 98/02354.

The endogenous murine olfactory receptor gene can be interrupted by the insertion, between two contiguous nucleotide of said gene, of a part of all of a marker gene placed under the control of the appropriate promoter, for example the endogenous promoter of the endogenous murine olfactory receptor gene. The marker gene may be the neomycin resistance gene (*neo*) that may be operably linked to the phosphoglycerate kinase-1 (PGK-1) promoter, as described in the PCT Application No WO 98/02534.

Thus, the invention is also directed to a transgenic animal contain in their somatic cells and/or in their germ line cells a polynucleotide selected from the following group of polynucleotides:

a) a defective copy of the human olfactory receptor gene;

b) a defective copy of the endogenous olfactory receptor gene, wherein the expression "endogenous olfactory receptor gene" designates an olfactory receptor gene that is naturally present within the genome of the animal host to be genetically modified.

The invention also concerns a method for obtaining transgenic animals, wherein said methods comprise the steps of :

a) replacing the endogenous copy of the animal olfactory receptor gene by a nucleic acid selected from the group consisting of a defective copy of the human olfactory receptor gene and a defective copy of the endogenous olfactory receptor gene in animal cells, preferably embryonic stem cells (ES);

b) introducing the recombinant animal cells obtained at step a) in embryos, notably blastocysts of the animal;

c) selecting the resulting transgenic animals, for example by detecting the defective copy of an olfactory receptor gene with one or several primers or probes according to the invention.

Optionally, the transgenic animals may be bred together in order to obtain homozygous transgenic animals for the defective copy of the olfactory receptor gene introduced.

The transgenic animals of the invention thus contain specific sequences of exogenous genetic material such as the nucleotide sequences described above in detail.

In a preferred embodiment, these transgenic animals may be good experimental models in order to study the diverse pathologies related to disorders associated to alteration of the olfactory

perception of odorant substances or molecules, in particular concerning the transgenic animals within the genome of which has been inserted one or several copies of a polynucleotide encoding a native olfactory receptor protein, or alternatively a mutant olfactory receptor protein.

Third preferred transgenic animals according to the invention contains in their somatic cells and/or in their germ line cells a polynucleotide selected from the following group of polynucleotides :

a) purified or isolated nucleic acid encoding an olfactory receptor polypeptide selected from OLF1 to OLF10, or a polypeptide fragment or variant thereof.

b) a purified or isolated nucleic comprising at least 8 consecutive nucleotides of the nucleotide sequence SEQ ID No 1, a nucleotide sequence complementary thereto or a fragment or a variant thereof;

c) a purified or isolated nucleic acid comprising a nucleotide sequence selected from the group of SEQ ID 2-11, a sequence complementary thereto or a fragment or a variant thereof.

The transgenic animals of the invention thus contain specific sequences of exogenous genetic material such as the nucleotide sequences described above in detail.

In a first preferred embodiment, these transgenic animals may be good experimental models in order to screen the candidate substance of interest interacting with the olfactory receptor under consideration.

Since it is possible to produce transgenic animals of the invention using a variety of different sequences, a general description will be given of the production of transgenic animals by referring generally to exogenous genetic material. This general description can be adapted by those skilled in the art in order to incorporate the DNA sequences into animals. For more details regarding the production of transgenic animals, and specifically transgenic mice, it may be referred to Sandou et al. (1994) and also to US Patents Nos 4,873,191, issued Oct.10, 1989, 5,968,766, issued Dec. 16, 1997 and 5,387,742, issued Feb. 28, 1995.

Transgenic animals of the present invention are produced by the application of procedures which result in an animal with a genome that incorporates exogenous genetic material which is integrated into the genome. The procedure involves obtaining the genetic material, or a portion thereof, which encodes either a coding sequence, a non-coding polynucleotide or a DNA sequence encoding an antisense polynucleotide of an olfactory receptor selected from the group OLF1 to OLF10 such as described in the present specification.

A recombinant polynucleotide of the invention is inserted into an embryonic or ES stem cell line. The insertion is made using electroporation. The cells subjected to electroporation are screened (e.g. Southern blot analysis) to find positive cells which have integrated the exogenous recombinant polynucleotide into their genome. An illustrative positive-negative selection procedure that may be used according to the invention is described by Mansour et al. (1988). Then, the positive cells are isolated, cloned and injected into 3.5 days old blastocysts from mice. The blastocysts are then

inserted into a female host animal and allowed to grow to term. The offsprings of the female host are tested to determine which animals are transgenic e.g. include the inserted exogenous DNA sequence and which are wild-type.

Thus, the present invention also concerns a transgenic animal containing a nucleic acid, a recombinant expression vector or a recombinant host cell according to the invention.

Recombinant Cell Lines Derived From The Transgenic Animals Of The Invention.

A further object of the invention comprises recombinant host cells obtained from a transgenic animal described herein. In one embodiment the invention encompasses cells derived from non-human host mammals and animals comprising a recombinant vector of the invention or an olfactory receptor gene disrupted by homologous recombination with a knock out vector.

Recombinant cell lines may be established *in vitro* from cells obtained from any tissue of a transgenic animal according to the invention, for example by transfection of primary cell cultures with vectors expressing *onc*-genes such as SV40 large T antigen, as described by Chou (1989) and Shay et al.(1991).

F. METHODS FOR SCREENING SUBSTANCES OR MOLECULES INTERACTING WITH AN OLFACTORY RECEPTOR PROTEIN

The present invention pertains to methods for screening substances of interest, in particular odorant substances or molecules that interact with an olfactory receptor protein selected from the group consisting of OLF1 to OLF10, or one peptide fragment or variant thereof. In one embodiment, the candidate substance is devoid of odorant propriety but it is able to bind the olfactory receptor and to trigger the transduction of signals.

For the purpose of the present invention, a ligand means a molecule, such as a protein, a peptide, an antibody or any synthetic chemical compound capable of binding to the olfactory receptor protein or one of its fragments or variants or to modulate the expression of the polynucleotide coding for olfactory receptor or a fragment or variant thereof.

In the ligand screening method according to the present invention, a biological sample or a defined molecule to be tested as a putative ligand of the olfactory receptor protein is brought into contact with the corresponding purified olfactory receptor protein, for example the corresponding purified recombinant olfactory receptor protein produced by a recombinant cell host as described herein, in order to form a complex between this protein and the putative ligand molecule to be tested.

As an illustrative example, to study the interaction of the olfactory receptor protein, or a fragment comprising comprising any of the fragments described in the section "OLF1 to OLF10 proteins and polypeptide fragments" with drugs or small molecules, such as molecules generated through combinatorial chemistry approaches, the microdialysis coupled to HPLC method described by Wang et al. (1997) or the affinity capillary electrophoresis method described by Bush et al. (1997) can be used.

In further methods, peptides, drugs, fatty acids, lipoproteins, or small molecules which interact with the olfactory receptor protein, or a fragment comprising any of the fragments described in the section "OLF1 to OLF10 proteins and polypeptide fragments" may be identified using assays such as the following. The molecule to be tested for binding is labeled with a detectable label, such as a fluorescent , radioactive, or enzymatic tag and placed in contact with immobilized olfactory receptor protein, or a fragment thereof under conditions which permit specific binding to occur, such as affinity columns. In some embodiments, chimeric proteins containing the olfactory receptor protein fused to proteins facilitating purification, such as glutathion S transferase (GST) are used. After removal of non-specifically bound molecules, bound molecules are detected using appropriate means.

In one embodiment, proteins, peptides, carbohydrates, lipids, or small molecules generated by combinatorial chemistry interacting with the olfactory receptor protein, or a fragment or a variant thereof can also be screened by using an Optical Biosensor as described in Edwards and Leatherbarrow (1997) and also in Szabo et al. (1995). The main advantage of the method is that it allows the determination of the association rate between the olfactory receptor protein and molecules interacting with the olfactory receptor protein. It is thus possible to select specifically ligand molecules interacting with the olfactory receptor protein, or a fragment thereof, through strong or conversely weak association constants.

Another object of the present invention comprises methods and kits for the screening of candidate substances that interact with olfactory receptor polypeptide.

The present invention pertains to methods for screening substances of interest that interact with an olfactory receptor protein or one fragment or variant thereof. By their capacity to bind covalently or non-covalently to an olfactory receptor protein or to a fragment or variant thereof, these substances or molecules may be advantageously used both *in vitro* and *in vivo*. *In vitro*, said interacting molecules may be used as detection means in order to identify the presence of an olfactory receptor protein in a sample, preferably a biological sample.

A first method for the screening of a candidate substance interacting with an olfactory receptor polypeptide selected from the group consisting of SEQ ID Nos 12-21, or fragments or variants thereof, comprises the following steps :

a) providing a polypeptide selected from the group consisting of the polypeptides comprising, consisting essentially of, or consisting of the amino acid sequences of SEQ ID Nos 12-21 , or a peptide fragment or a variant thereof;

b) obtaining a candidate substance;

c) bringing into contact said polypeptide with said candidate substance; and

d) detecting the complexes formed between said polypeptide and said candidate substance.

Various candidate substances or molecules can be assayed for interaction with an olfactory receptor polypeptide. These substances or molecules include, without being limited to, natural or synthetic organic compounds or molecules of biological origin such as polypeptides. When the candidate substance or molecule comprises a polypeptide, this polypeptide may be the resulting expression product of either a phage clone belonging to a phage-based random peptide library, or of a cDNA library cloned in a vector suitable for performing a two-hybrid screening assay.

In one embodiment of the screening method defined above, the complexes formed between the polypeptide and the candidate substance are further incubated in the presence of a polyclonal or a monoclonal antibody that specifically binds to the olfactory receptor protein of the invention under consideration or to said peptide fragment or variant thereof.

In another embodiment of the present screening method, increasing concentrations of a substance competing for binding to the olfactory receptor with the considered candidate substance is added, simultaneously or prior to the addition of the candidate substance or molecule, when performing step c) of said method. By this technique, the detection and optionally the quantification of the complexes formed between the olfactory receptor protein or the peptide fragment or variant thereof and the candidate substance or molecule to be screened allows the one skilled in the art to determine the affinity value of said substance or molecule for said olfactory receptor protein or the peptide fragment or variant thereof.

The olfactory receptor selected from the group consisting of OLF1 to OLF10, or a peptide fragment or a variant thereof, can be overexpressed and purified in a bacterial system such as E coli as described in Kiefer et al. (1996) and Tucker et al. (1996). The olfactory receptor coding sequence can be fused to its N-terminus with GST (Glutathione S transferase) or MBP (Maltose Binding Protein) and to its C-terminus with poly-histidine tag, Bio tag or Strep tag for facilitating the purification of the expressed protein. The Bio tag is 13 amino acid residues long, is biotinylated *in vivo* in E. coli, and will therefore bind to both avidin and streptavidin. The Strep tag is 9 amino acid residues long and binds specifically to streptavidin, but not to avidin. Therewith, a purification step by affinity can be carried out based on the interaction of a poly-histidine tail with immobilized metal ions, of the biotinylated Bio tag with monomeric avidin, of the Strep tag with streptavidin, of the GST segment with the glutathione, or of the MBP segment with the maltose. Thioredoxin can be eventually inserted between the receptor C-terminus and the tag and could increase the expression level. The fusion protein is solubilized in 1% N-laurroyl sarcosine, and 0.2 % digitonin is added. It is purified by affinity chromatography. The MBP, GST or tag segment can be then removed. After the olfactory receptor protein purification, sarcosyl can be replaced with digitonin which is a detergent widely used to stabilize the G protein-coupled receptors. The purified receptor is reconstituted into lipid vesicles preferably composed of phosphatidylcholine: phosphatidylglycerol (4:1) by adding the lipid dissolved in dodecyl maltoside and removing the detergent.

The olfactory receptor selected from the group consisting of OLF1 to OLF10, or a peptide fragment or a variant thereof, can also be overexpressed and purified in a baculovirus/Sf9 system as described in Nekrasova et al. (1996). The olfactory receptor gene, or a fragment thereof, is preferably expressed with a "flag" peptide epitope tag and/or a poly-histidine tag to either its N- or C-terminus for facilitating the purification of the expressed protein. Therefore, the olfactory receptor gene, or a fragment or a variant thereof, is preferably subcloned into the baculovirus transfer vector pAcSGHisNT to create constructs that encoded olfactory receptor with amino-terminal poly-histidine tag. The resulting transfer vector is transfected preferably with BaculoGold DNA into Sf9 cells. The expressed olfactory receptors are then solubilized either in 1 % N-lauryl sarcosine or 1.5 % lysophosphatidylcholine, but preferably in lysophosphatidylcholine. After solubilization, the olfactory receptors are purified by affinity chromatography on nickel nitrilotriacetic acid resin and by cation-exchange chromatography with carboxymethyl sepharaose cation-exchange column. The tag segment can be then removed. The purified receptor is reconstituted into lipid vesicles preferably composed of dimyritoylglycerophosphocholine, cholesterol, dialmitoylgycerophosphoserine and dipalmitoylglycerophosphoethanolamine (in molecular ratio 54:35:10:1)

Once the olfactory receptor protein or one of its peptide fragments or variants has been obtained as described above, candidate substances or molecules can then be assayed for their capacity to bind thereto.

The candidate substance or molecule to be assayed for interacting with an olfactory receptor of the invention may be of diverse nature, including, without being limited to, natural or synthetic organic compounds or molecules of biological origin such as peptide. It can comprise aromatic or aliphatic compounds with various functional groups such as alcohol, aldehyde, ester, ether, ketone, carboxylic, amine. An example of a substance panel which can be used is provided by Zhao et al. (1998).

The screening of substances or molecules interacting with an olfactory receptor, or a fragment thereof, is carried out by photoaffinity labeling experiments described in Kiefer et al. (1996). The odorant is labeled, preferably radiolabeled, and incubated with lipid vesicles including the purified olfactory receptor. The odorants bound to the olfactory receptors are crosslinked by exposure to ultraviolet light. Then, the samples are subjected to SDS polyacrylamide gel electrophoresis. Proteins are visualized by Coomassie-blue staining and the odorants are revealed, preferably by autoradiography. In another embodiment, the proteins can be visualized by Western Blot with a polyclonal or monoclonal antibody that specifically binds to the olfactory receptor under consideration. Once a substance binding to the considered olfactory receptor is identified, the binding specificity of this substance is confirmed with competition experiments demonstrating that increasing concentrations of unlabeled ligand accomplish a dose-dependent displacement of the radioactive ligand.

The identification of a first substance specific to one of the olfactory receptors of the present invention facilitates the screening of other substances. Indeed, the binding capacity of the screened substances to this olfactory receptor can be carried out through a competition experiments against the first identified substance which is labeled.

The invention also pertains to kits useful for performing the hereinbefore described screening method. Preferably, such kits comprise a polypeptide selected form the group consisting of the polypeptides comprising the amino acid sequences SEQ ID Nos 12-21 or a peptide fragment or a variant thereof, and optionally means useful to detect the complex formed between the considered olfactory receptor polypeptide or its peptide fragment or variant and the candidate substance. In a preferred embodiment, the kit can comprise an already identified substance specific of the olfactory receptor under consideration which is labeled, preferably radiolabeled, and a monoclonal or polyclonal antibody directed against the considered olfactory receptor.

A second screening method embodiment consists of a method for the screening of ligand molecules interacting with an olfactory receptor polypeptide selected from the group consisting of SEQ ID Nos 12-21, wherein said method comprises :

a) providing a recombinant eukaryotic host cell containing a nucleic acid encoding a polypeptide selected from the group comprising, consisting essentially of, or consisting the polypeptides comprising the amino acid sequences SEQ ID Nos 12-21, or variants or fragments thereof;

b) preparing membrane extracts of said recombinant eukaryotic host cell;

c) bringing into contact the membrane extracts prepared at step b) with a selected ligand molecule; and

d) detecting the production level of second messengers metabolites.

The baculovirus-Sf9 cell system enables a foreign DNA encoding an olfactory receptor selected from the group consisting of OLF1 to OLF10, or a peptide fragment or a variant thereof, to be expressed with high efficiency. Moreover, it can be used to couple a heterologous expressed olfactory receptor to the second messenger cascades. Therefore, the binding specificity of an olfactory receptor can be assessed through an assay of odorant-induced generation of cAMP or inositol triphosphate (InsP3) described in Raming et al. (1993).

Briefly, a cell line derived from Sf9 is infected by baculovirus, such as baculovirus transfer vector pVL1393, harboring DNA encoding the olfactory receptor or a fragment thereof downstream from a strong promoter, preferably the polyhedrin promoter. Recombinant virus are purified and used to infect 1.5×10^8 Sf9 cells in 100 ml spinner cultures at high multiplicity of infection. Cells are collected after a postinfection delay, preferably 48 h, and membrane fractions are isolated as follow.

Cells are pelleted (at 250g for 10 min at 4°C), washed with Ringer solution (120 mM NaCl, 5 mM KCl, 1.6 mM K_2HPO_4, 1.2 mM $MgSO_4$, 25 mM $NaHCO_3$, 5 mM glucose, pH7.4) and disrupted using a glass homogenizer in homogenization buffer (10 mM Tris-HCl, pH 8.0, 2 mM

EGTA, 3 mM MgCl2) containing antiproteases. The homogenate is centrifuged and the pellet is washed. Supernatants are centrifuged at 33,000g for 20 min. The final pellet is resuspended in homogenization buffer and the protein concentration is determined.

Assay of odorant substance-induced generation of second messengers cAMP and InsP3 is performed as follow. Suspensions of Sf9 cell membrane preparations (300 µg protein) are rapidly mixed with a stimulation buffer (200 mM NaCl, 10 mM EGTA, 50 mM MOPS, 2.5 mM $MgCl_2$, 1 mM DTT, 0.05 % Na-cholate, 1 mM ATP, 1 µM GTP, and 0.02 µM free Ca^{2+}) containing the candidate substances at the appropriate concentrations. The reaction is stopped after a short time, preferably 1 sec, by injecting 10 % Perchloric acid. Quenched samples are assayed for second messenger concentrations. The quenched and cooled samples are vortexed and centrifuged for 5 min at 2500g at 4°C. 400 µl of the supernatants are transferred to a separate tube containing 100 µl of 10 mM EDTA (pH 7). The sample are neutralized by adding 500 µl of a 1:1 (v/v) mixture of 1,1,2 trichlorofluoroethane, followed by thorough mixing. After centrifugation for 2 min at 500g, three phases are obtained. The upper phase, which contains all water soluble components, is used for carrying out the concentration measurements. cAMP and InsP3 concentrations are determined according the procedure of Steiner et al. (1972) and Palmer et al. (1989), respectively.

The invention also concerns a kit for the screening of odorant ligand molecules interacting with an olfactory receptor polypeptide selected from the group consisting of the polypeptides comprising the amino acid sequences SEQ ID Nos 12-21, wherein said kit comprises :

a) a recombinant eukaryotic host cell containing a nucleic acid encoding a polypeptide selected from the group comprising, consisting essentially of, or consisting of the polypeptides comprising the amino acid sequences SEQ ID Nos 12-21 or variants or fragments thereof; and

b) optionally, reagents necessary for the measurement of second messenger metabolites in a sample.

The screening of substances or molecules interacting with an olfactory receptor, or a fragment thereof, can also be carried out through the measurement of the increase of the response to odorants in an olfactory epithelium overexpressing an olfactory receptor selected from the group consisting of OLF1 to OLF10, or a peptide fragment or a variant thereof, as described in Zhao et al. (1998). The response is assessed by electro-olfactogram which measures a transepithelial potential resulting from the summed activity of many olfactory neurons. In order to overexpress the olfactory receptor, or a fragment thereof, in an olfactory epithelium, an adenovirus containing the olfactory receptor gene is generated. To aid in electro-olfactogram electrode placements, the olfactory receptor coding sequence is preferably combined in the adenovirus with the physiological marker green fluorescent protein (GFP) in such manner that the two proteins are simultaneously expressed. The olfactory epithelium of an animal, preferably of a rat, is infected by the adenovirus. Animals are killed 3 to 8 days after infection and the nasal cavity is opened, exposing the medial surface of the

nasal turbinates. Under fluorescent illumination, the GFP clearly marked the pattern of viral infection and olfactory receptor expression. Odorant substance are applied to the olfactory epithelium in the vapor phase by injecting a pressurized pulse of odorant vapor into a continuous stream of humidified clean air. Electro-olfactogram recordings are obtained with a glass capillary electrode placed on the surface of the epithelium and connected to a differential amplifier. The olfactory receptor specificity is assessed from the increase of response in infected animals compared to uninfected animals. To account for the variability between animals, a standard odorant to which all other odorant responses are normalized is used.

A third screening method embodiment consists of a method for the screening of ligand molecules interacting with an olfactory receptor polypeptide selected from the group consisting of the polypeptides comprising the amino acid sequences SEQ ID Nos 12-21, wherein said method comprises :

a) providing an adenovirus containing a nucleic acid encoding a polypeptide selected from the group comprising, consisting essentially of, or consisting of the polypeptides comprising the amino acid sequences SEQ ID Nos 12-21, or variants or fragments thereof;

b) infecting an olfactory epithelium with said adenovirus;

c) bringing into contact the olfactory epithelium b) with a selected ligand molecule; and

d) detecting the increase of the response to said ligand molecule.

G. METHODS FOR INHIBITING THE EXPRESSION OF AN OLFACTORY RECEPTOR GENE

Other therapeutic compositions according to the present invention comprise advantageously an oligonucleotide fragment of the nucleic sequence of olfactory receptor as an antisense tool or a triple helix tool that inhibits the expression of the corresponding olfactory receptor gene. A preferred fragment of the nucleic sequence of olfactory receptor comprises an allele of at least one of the biallelic markers A1 to A13.

Antisense Approach

Preferred methods using antisense polynucleotide according to the present invention are the procedures described by Sczakiel et al.(1995).

Preferred antisense polynucleotides are described in the section entitled "Nuclear Antisense DNA Constructs".

The antisense nucleic acids should have a length and melting temperature sufficient to permit formation of an intracellular duplex having sufficient stability to inhibit the expression of the olfactory receptor mRNA in the duplex. Strategies for designing antisense nucleic acids suitable for use in gene therapy are disclosed in Green et al., (1986) and Izant and Weintraub, (1984).

In some strategies, antisense m lecules are obtained by reversing the orientation of the olfactory receptor coding region with respect to a promoter so as to transcribe the opposite strand from that which is normally transcribed in the cell. The antisense molecules may be transcribed using in vitro transcription systems such as those which employ T7 or SP6 polymerase to generate the transcript. Another approach involves transcription of olfactory receptor antisense nucleic acids in vivo by operably linking DNA containing the antisense sequence to a promoter in a suitable expression vector.

Alternatively, suitable antisense strategies are those described by Rossi et al.(1991), in the International Applications Nos. WO 94/23026, WO 95/04141, WO 92/18522 and in the European Patent Application No. EP 0 572 287 A2

An alternative to the antisense technology that is used according to the present invention comprises using ribozymes that will bind to a target sequence via their complementary polynucleotide tail and that will cleave the corresponding RNA by hydrolyzing its target site (namely "hammerhead ribozymes"). Briefly, the simplified cycle of a hammerhead ribozyme comprises (1) sequence specific binding to the target RNA via complementary antisense sequences; (2) site-specific hydrolysis of the cleavable motif of the target strand; and (3) release of cleavage products, which gives rise to another catalytic cycle. Indeed, the use of long-chain antisense polynucleotide (at least 30 bases long) or ribozymes with long antisense arms are advantageous. A preferred delivery system for antisense ribozyme is achieved by covalently linking these antisense ribozymes to lipophilic groups or to use liposomes as a convenient vector. Preferred antisense ribozymes according to the present invention are prepared as described by Sczakiel et al.(1995), the specific preparation procedures being referred to in said article.

Triple Helix Approach

The olfactory receptor genomic DNA may also be used to inhibit the expression of the olfactory receptor gene based on intracellular triple helix formation.

Triple helix oligonucleotides are used to inhibit transcription from a genome. They are particularly useful for studying alterations in cell activity when it is associated with a particular gene.

Similarly, a portion of the olfactory receptor genomic DNA can be used to study the effect of inhibiting olfactory receptor transcription within a cell. Traditionally, homopurine sequences were considered the most useful for triple helix strategies. However, homopyrimidine sequences can also inhibit gene expression. Such homopyrimidine oligonucleotides bind to the major groove at homopurine:homopyrimidine sequences. Thus, both types of sequences from the olfactory receptor gen mic DNA are contemplated within the scope of this invention.

To carry out gene therapy strategies using the triple helix approach, the sequences of the olfactory receptor genomic DNA are first scanned to identify 10-mer to 20-mer homopyrimidine or homopurine stretches which could be used in triple-helix based strategies for inhibiting olfactory

receptor expression. Following identification of candidate homopyrimidine or homopurine stretches, their efficiency in inhibiting olfactory receptor expression is assessed by introducing varying amounts of oligonucleotides containing the candidate sequences into tissue culture cells which express the olfactory receptor gene.

The oligonucleotides can be introduced into the cells using a variety of methods known to those skilled in the art, including but not limited to calcium phosphate precipitation, DEAE-Dextran, electroporation, liposome-mediated transfection or native uptake.

Treated cells are monitored for altered cell function or reduced olfactory receptor expression using techniques such as Northern blotting, RNase protection assays, or PCR based strategies to monitor the transcription levels of the olfactory receptor gene in cells which have been treated with the oligonucleotide.

The oligonucleotides which are effective in inhibiting gene expression in tissue culture cells may then be introduced in vivo using the techniques described above in the antisense approach at a dosage calculated based on the in vitro results, as described in antisense approach.

In some embodiments, the natural (beta) anomers of the oligonucleotide units can be replaced with alpha anomers to render the oligonucleotide more resistant to nucleases. Further, an intercalating agent such as ethidium bromide, or the like, can be attached to the 3′ end of the alpha oligonucleotide to stabilize the triple helix. For information on the generation of oligonucleotides suitable for triple helix formation see Griffin et al.(1989).

H. COMPUTER-RELATED EMBODIMENTS

As used herein the term "nucleic acid codes of the invention" encompass the nucleotide sequences comprising, consisting essentially of, or consisting of any of the polynucleotides described in the "Coding Regions of the olfactory receptor gene" section, "Genomic sequence of the olfactory receptor gene" section and the "Oligonucleotide Probes And Primers" section, or variants thereof, or complementary sequences thereto. Homologous sequences refer to a sequence having at least 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, or 75% homology to these contiguous spans. Homology may be determined using any method described herein, including BLAST2N with the default parameters or with any modified parameters. Homologous sequences also may include RNA sequences in which uridines replace the thymines in the nucleic acid codes of the invention.

As used herein the term "polypeptide codes of the invention" encompass the polypeptide sequences comprising any of the polypeptides described in the " OLF1 to OFL10 proteins and polypeptide fragments".

It will be appreciated that the nucleic acid and polypeptide codes of the invention can be represented in the traditional single character format or three letter format respectively (See the inside back cover of Stryer, Lubert. *Biochemistry*, 3rd edition. W. H Freeman & Co., New York.) or in any

other format or code which records the identity of the nucleotides or the amino acid respectively in a sequence.

It will be appreciated by those skilled in the art that the nucleic acid codes of the invention and polypeptide codes of the invention can be stored, recorded, and manipulated on any medium which can be read and accessed by a computer. As used herein, the words "recorded" and "stored" refer to a process for storing information on a computer medium. A skilled artisan can readily adopt any of the presently known methods for recording information on a computer readable medium to generate manufactures comprising one or more of the nucleic acid codes of the invention, or one or more of the polypeptide codes of the invention. Another aspect of the present invention is a computer readable medium having recorded thereon at least 2, 5, 10, 15, 20, 25, 30, or 50 nucleic acid codes of the invention. Another aspect of the present invention is a computer readable medium having recorded thereon at least 2, 5, 10, 15, 20, 25, 30, or 50 polypeptide codes of the invention.

Computer readable media include magnetically readable media, optically readable media, electronically readable media and magnetic/optical media. For example, the computer readable media may be a hard disc, a floppy disc, a magnetic tape, CD-ROM, DVD, RAM, or ROM as well as other types of other media known to those skilled in the art.

Embodiments of the present invention include systems, particularly computer systems which contain the sequence information described herein. As used herein, "a computer system" refers to the hardware components, software components, and data storage components used to store and/or analyze the nucleotide sequences of the nucleic acid codes of the invention, the amino acid sequences of the polypeptide codes of the invention, or other sequences. The computer system preferably includes the computer readable media described above, and a processor for accessing and manipulating the sequence data.

In some embodiments, the computer system may further comprise a sequence comparer for comparing the nucleic acid codes or polypeptide codes of the invention stored on a computer readable medium to reference nucleotide sequences stored on a computer readable medium. A "sequence comparer" refers to one or more programs which are implemented on the computer system to compare a nucleotide or polypeptide sequence with other nucleotide or polypeptide sequences and/or compounds including but not limited to peptides, peptidomimetics, and chemicals the sequences or structures of which are stored within the data storage means. For example, the sequence comparer may compare the nucleotide sequences of the nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention stored on a computer readable medium to reference sequences stored on a computer readable medium to identify homologies, motifs implicated in biological function, or structural motifs. The various sequence comparer programs identified elsewhere in this patent specification are particularly contemplated for use in this aspect of the invention.

Accordingly, one aspect of the present invention is a computer system comprising a processor, a data storage device having stored thereon a nucleic acid code of the invention or a

polypeptide code of the invention, a data storage device having retrievably stored thereon reference nucleotide sequences or polypeptide sequences to be compared to the nucleic acid code of the invention or polypeptide code of the invention and a sequence comparer for conducting the comparison. The sequence comparer may indicate a homology level between the sequences compared or identify structural motifs in the nucleic acid code of the invention and polypeptide codes of the invention or it may identify structural motifs in sequences which are compared to these nucleic acid codes and polypeptide codes. In some embodiments, the data storage device may have stored thereon the sequences of at least 2, 5, 10, 15, 20, 25, 30, or 50 of the nucleic acid codes of the invention or polypeptide codes of the invention.

Another aspect of the present invention is a method for determining the level of homology between a nucleic acid code of the invention and a reference nucleotide sequence, comprising the steps of reading the nucleic acid code and the reference nucleotide sequence through the use of a computer program which determines homology levels and determining homology between the nucleic acid code and the reference nucleotide sequence with the computer program. The computer program may be any of a number of computer programs for determining homology levels, including those specifically enumerated herein, including BLAST2N with the default parameters or with any modified parameters. The method may be implemented using the computer systems described above. The method may also be performed by reading 2, 5, 10, 15, 20, 25, 30, or 50 of the above described nucleic acid codes of the invention through the use of the computer program and determining homology between the nucleic acid codes and reference nucleotide sequences.

Alternatively, the computer program may be a computer program which compares the nucleotide sequences of the nucleic acid codes of the present invention, to reference nucleotide sequences in order to determine whether the nucleic acid code of the invention differs from a reference nucleic acid sequence at one or more positions. Optionally such a program records the length and identity of inserted, deleted or substituted nucleotides with respect to the sequence of either the reference polynucleotide or the nucleic acid code of the invention. In one embodiment, the computer program may be a program which determines whether the nucleotide sequences of the nucleic acid codes of the invention contain one or more single nucleotide polymorphisms (SNP) with respect to a reference nucleotide sequence. These single nucleotide polymorphisms may each comprise a single base substitution, insertion, or deletion.

Another aspect of the present invention is a method for determining the level of homology between a polypeptide code of the invention and a reference polypeptide sequence, comprising the steps of reading the polypeptide code of the invention and the reference polypeptide sequence through use of a computer program which determines homology levels and determining homology between the polypeptide code and the reference polypeptide sequence using the computer program.

Accordingly, another aspect of the present invention is a method for determining whether a nucleic acid code of the invention differs at one or more nucleotides from a reference nucleotide

sequence comprising the steps of reading the nucleic acid code and the reference nucleotide sequence through use of a computer program which identifies differences between nucleic acid sequences and identifying differences between the nucleic acid code and the reference nucleotide sequence with the computer program. In some embodiments, the computer program is a program which identifies single nucleotide polymorphisms. The method may be implemented by the computer systems described above. The method may also be performed by reading at least 2, 5, 10, 15, 20, 25, 30, or 50 of the nucleic acid codes of the invention and the reference nucleotide sequences through the use of the computer program and identifying differences between the nucleic acid codes and the reference nucleotide sequences with the computer program.

An "identifier" refers to one or more programs which identifies certain features within the above-described nucleotide sequences of the nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention.

In one embodiment, the identifier may comprise a molecular modeling program which determines the 3-dimensional structure of the polypeptides codes of the invention. In some embodiments, the molecular modeling program identifies target sequences that are most compatible with profiles representing the structural environments of the residues in known three-dimensional protein structures. (See, e.g., Eisenberg et al., U.S. Patent No. 5,436,850 issued July 25, 1995). In another technique, the known three-dimensional structures of proteins in a given family are superimposed to define the structurally conserved regions in that family. This protein modeling technique also uses the known three-dimensional structure of a homologous protein to approximate the structure of the polypeptide codes of the invention. (See e.g., Srinivasan, et al., U.S. Patent No. 5,557,535 issued September 17, 1996). Conventional homology modeling techniques have been used routinely to build models of proteases and antibodies. (Sowdhamini et al., (1997)). Comparative approaches can also be used to develop three-dimensional protein models when the protein of interest has poor sequence identity to template proteins. In some cases, proteins fold into similar three-dimensional structures despite having very weak sequence identities. For example, the three-dimensional structures of a number of helical cytokines fold in similar three-dimensional topology in spite of weak sequence homology.

The recent development of threading methods now enables the identification of likely folding patterns in a number of situations where the structural relatedness between target and template(s) is not detectable at the sequence level. Hybrid methods, in which fold recognition is performed using Multiple Sequence Threading (MST), structural equivalencies are deduced from the threading output using a distance geometry program DRAGON to construct a low resolution model, and a full-atom representation is constructed using a molecular modeling package such as QUANTA. According to this 3-step approach, candidate templates are first identified by using the novel fold recognition algorithm MST, which is capable of performing simultaneous threading of multiple aligned sequences onto one or more 3-D structures. In a second step, the structural

equivalencies obtained from the MST output are converted into interresidue distance restraints and fed into the distance geometry program DRAGON, together with auxiliary information obtained from secondary structure predictions. The program combines the restraints in an unbiased manner and rapidly generates a large number of low resolution model confirmations. In a third step, these low resolution model confirmations are converted into full-atom models and subjected to energy minimization using the molecular modeling package QUANTA. (See e.g., Aszódi et al., (1997)).

he results of the molecular modeling analysis may then be used in rational drug design techniques to identify agents which modulate the activity of the polypeptide codes of the invention.

Accordingly, another aspect of the present invention is a method of identifying a feature within the nucleic acid codes of the invention or the polypeptide codes of the invention comprising reading the nucleic acid code(s) or the polypeptide code(s) through the use of a computer program which identifies features therein and identifying features within the nucleic acid code(s) or polypeptide code(s) with the computer program. In one embodiment, computer program comprises a computer program which identifies open reading frames. In a further embodiment, the computer program identifies structural motifs in a polypeptide sequence. In another embodiment, the computer program comprises a molecular modeling program. The method may be performed by reading a single sequence or at least 2, 5, 10, 15, 20, 25, 30, or 50 of the nucleic acid codes of the invention or the polypeptide codes of the invention through the use of the computer program and identifying features within the nucleic acid codes or polypeptide codes with the computer program.

The nucleic acid codes of the invention or the polypeptide codes of the invention may be stored and manipulated in a variety of data processor programs in a variety of formats. For example, they may be stored as text in a word processing file, such as MicrosoftWORD or WORDPERFECT or as an ASCII file in a variety of database programs familiar to those of skill in the art, such as DB2, SYBASE, or ORACLE. In addition, many computer programs and databases may be used as sequence comparers, identifiers, or sources of reference nucleotide or polypeptide sequences to be compared to the nucleic acid codes of the invention or the polypeptide codes of the invention. The following list is intended not to limit the invention but to provide guidance to programs and databases which are useful with the nucleic acid codes of the invention or the polypeptide codes of the invention. The programs and databases which may be used include, but are not limited to: MacPattern (EMBL), DiscoveryBase (Molecular Applications Group), GeneMine (Molecular Applications Group), Look (Molecular Applications Group), MacLook (Molecular Applications Group), BLAST and BLAST2 (NCBI), BLASTN and BLASTX (Altschul et al, 1990), FASTA (Pearson and Lipman, 1988), FASTDB (Brutlag et al., 1990), Catalyst (Molecular Simulations Inc.), Catalyst/SHAPE (Molecular Simulations Inc.), Cerius^2.DBAccess (Molecular Simulations Inc.), HypoGen (Molecular Simulations Inc.), Insight II, (Molecular Simulations Inc.), Discover (Molecular Simulations Inc.), CHARMm (Molecular Simulations Inc.), Felix (Molecular Simulations Inc.), DelPhi, (Molecular Simulations Inc.), QuanteMM, (Molecular Simulations Inc.), Homology (Molecular Simulations Inc.), Modeler

(Molecular Simulations Inc.), ISIS (Molecular Simulations Inc.), Quanta/Protein Design (Molecular Simulations Inc.), WebLab (Molecular Simulations Inc.), WebLab Diversity Explorer (Molecular Simulations Inc.), Gene Explorer (Molecular Simulations Inc.), SeqFold (Molecular Simulations Inc.), the EMBL/Swissprotein database, the MDL Available Chemicals Directory database, the MDL Drug Data Report data base, the Comprehensive Medicinal Chemistry database, Derwents's World Drug Index database, the BioByteMasterFile database, the Genbank database, and the Genseqn database. Many other programs and data bases would be apparent to one of skill in the art given the present disclosure.

Motifs which may be detected using the above programs include sequences encoding leucine zippers, helix-turn-helix motifs, glycosylation sites, ubiquitination sites, alpha helices, and beta sheets, signal sequences encoding signal peptides which direct the secretion of the encoded proteins, sequences implicated in transcription regulation such as homeoboxes, acidic stretches, enzymatic active sites, substrate binding sites, and enzymatic cleavage sites.

Throughout this application, various publications, patents and published patent applications are cited. The disclosures of these publications, patents and published patent specification referenced in this application are hereby incorporated by reference into the present disclosure to more fully describe the sate of the art to which this invention pertains.

EXAMPLES

EXAMPLE 1: LOCALIZATION OF THE OLFACTORY RECEPTOR GENE OLF3 AND OLF5 ON THE HUMAN CHROMOSOMES.

Metaphase chromosome preparation

Metaphase chromosomes were prepared from phytohemagglutinin (PHA)-stimulated blood cell donors. PHA stimulated lymphocytes from healthy males were cultured for 72 h in RPMI-1640 medium. For synchronization, methotrexate (10 μM) was added for 17 h, followed by addition of 5-bromodeoxyuridine (5-BrdU, 0.1 mM) for 6 h. Colcemid (1 mg/ml) was added for the last 15 min before harvesting the cells. Cells were collected, washed in RPMI, incubated with a hypotonic solution of KCl (75 mM) at 37°C for 15 min and fixed in three changes of methanol:acid acetic (3:1). The cell suspension was dropped onto a glass slide, air-dried and kept in darkness at –20°C until use.

Probes:

- The BAC H0526H04 containing Olf3 and Olf5 genes was used to generate probe by Alu-PCR. PCR amplification of BAC recombinant DNA (50 ng) was carried out as described by Romana et al. (1993).

- Two DNA fragments carrying respectively Olf3 and Olf5 sequences were generated by long range PCR with specific primers (SEQ ID 96-99) and used as probes to confirm the localization of each genes. Olf3 and Olf5 amplicons are respectively 2.8 kb and 3.2 kb fragments.

Probes were labeled by nick translation with bio-16-dUTP (Boehringer Mannheim), and purified over a Sephadex G50 column.

Fluorescence In Situ Hybridization

To determine the chromosomal localization of both genes, the BAC probe was initially hybridized to human metaphase cells. When biotinylated PCR products of BAC DNA were used in hybridization experiment, 75 ng of probe was precipitated with 75 µg of competitor DNA (human Cot1 DNA, GIBCO-BRL) and resuspended in 10 µl of hybridization buffer (50% formamide, 2 X SSC, 10% dextran sulfate, 1 mg/ml sonicated herring DNA, pH 7). When long range PCR products of Olf3 or Olf5 genes were used as probe, 5 ng of biotinylated probe were mixed with 5 µg of human Cot1 DNA. Prior to hybridization, the probe was denatured at 70°C for 10 min and preannealed at 37°C for 2 h.

Slides were treated for 1 h at 37°C with Rnase A (100 µg/ml), rinsed three times in 2 X SSC and dehydrated in an ethanol serie. Chromosome preparations were denatured in 70% formamide, 2 X SSC (pH 7), for 2 min at 70°C, then dehydrated at 4°C. The slides were treated with proteinase K (10 µg/ml in 20 mM Tris-HCl, 2 mM CaCl2) at 37°C for 8-10 min and dehydrated. After preannealing, the hybridization mixture containing the probe was placed on the slide, covered with a coverslip, sealed with rubber cement and incubated overnight in a humid chamber at 37°C. After hybridization and post hybridization washes, the biotinylated probe was detected by avidin-FITC (5 µg/ml, Vector Laboratories) and amplified once with additional layers of biotinylated goat anti-avidin (5 µg/ml, Vector Laboratories) and avidin-FITC. For chromosomal localization, fluorescent R-Bands were obtained as described by Cherif et al. (1990). The slides were observed under a LEICA fluorescent microscope (DMRXA). Chromosomes were counterstained with propidium iodide and the fluorescent signal of the probe appeared as two symmetrical yellow-green spots on both chromatids of the fluorescent R-band chromosome.

Localization

A specific signal (a double yellow-green spot) was observed on band 11q12-q13 on at least on chromosome 11 in >80% of the metaphases with all the probes.

EXAMPLE 2 : IDENTIFICATION OF BIALLELIC MARKERS: DNA EXTRACTION

Donors were unrelated and healthy. They presented a sufficient diversity for being representative of a French heterogeneous population. The DNA from 100 individuals was extracted and tested for the detection of the biallelic markers.

30 ml of peripheral venous blood were taken from each donor in the presence of EDTA. Cells (pellet) were collected after centrifugation for 10 minutes at 2000 rpm. Red cells were lysed by a lysis solution (50 ml final volume : 10 mM Tris pH7.6; 5 mM $MgCl_2$; 10 mM NaCl). The solution was centrifuged (10 minutes, 2000 rpm) as many times as necessary to eliminate the residual red cells present in the supernatant, after resuspension of the pellet in the lysis solution.

The pellet of white cells was lysed overnight at 42°C with 3.7 ml of lysis solution composed of:

- 3 ml TE 10-2 (Tris-HCl 10 mM, EDTA 2 mM) / NaCl 0.4 M
- 200 µl SDS 10%
- 500 µl K-proteinase (2 mg K-proteinase in TE 10-2 / NaCl 0.4 M).

For the extraction of proteins, 1 ml saturated NaCl (6M) (1/3.5 v/v) was added. After vigorous agitation, the solution was centrifuged for 20 minutes at 10000 rpm.

For the precipitation of DNA, 2 to 3 volumes of 100% ethanol were added to the previous supernatant, and the solution was centrifuged for 30 minutes at 2000 rpm. The DNA solution was rinsed three times with 70% ethanol to eliminate salts, and centrifuged for 20 minutes at 2000 rpm. The pellet was dried at 37°C, and resuspended in 1 ml TE 10-1 or 1 ml water. The DNA concentration was evaluated by measuring the OD at 260 nm (1 unit OD = 50 µg/ml DNA).

To determine the presence of proteins in the DNA solution, the OD 260 / OD 280 ratio was determined. Only DNA preparations having a OD 260 / OD 280 ratio between 1.8 and 2 were used in the subsequent examples described below.

The pool was constituted by mixing equivalent quantities of DNA from each individual.

EXAMPLE 3 : IDENTIFICATION OF BIALLELIC MARKERS: AMPLIFICATION OF GENOMIC DNA BY PCR

The amplification of specific genomic sequences of the DNA samples of example 2 was carried out on the pool of DNA obtained previously. In addition, 50 individual samples were similarly amplified.

PCR assays were performed using the following protocol:

Final volume	25 µl
DNA	2 ng/µl
$MgCl_2$	2 mM
dNTP (each)	200 µM
primer (each)	2.9 ng/µl
Ampli Taq Gold DNA polymerase	0.05 unit/µl
PCR buffer (10x = 0.1 M TrisHCl pH8.3 0.5M KCl	1x

Each pair of first primers was designed using the sequence information of the olfactory receptor gene cluster disclosed herein and the OSP software (Hillier & Green, 1991). This first pair

of primers was about 20 nucleotides in length and had the sequences disclosed in Table 1 in the columns labeled PU and RP.

Table 1

Amplicon	Position range of the amplicon in SEQ ID 1		Primer name RP	Position range of amplification primer in SEQ ID No 1		Primer name PU	Complementary position range of amplification primer in SEQ ID No 1	
99-13670	7362	7824	B1	7362	7380	C1	7805	7824
99-13669	8120	8662	B2	8120	8140	C2	8643	8662
99-13666	14308	14757	B3	14308	14328	C3	14740	14757
99-13664	19346	19845	B4	19346	19366	C4	19826	19845
99-13663	20298	20800	B5	20298	20318	C5	20781	20800
99-13660	76752	77223	B6	76752	76772	C6	77205	77223
99-13652	90967	91494	B7	90967	90987	C7	91474	91494
99-13671	133925	134393	B8	133925	133945	C8	134375	134393
99-13649	139807	140351	B9	139807	139826	C9	140331	140351
99-13648	140912	141434	B10	140912	140932	C10	141416	141434
99-13647	143828	144309	B11	143828	143847	C11	144292	144309

Preferably, the primers contained a common oligonucleotide tail upstream of the specific bases targeted for amplification which was useful for sequencing.

Primers PU contain the following additional PU 5' sequence : TGTAAAACGACGGCCAGT; primers RP contain the following RP 5' sequence : CAGGAAACAGCTATGACC. The primer containing the additional PU 5' sequence is listed in SEQ ID No 26. The primer containing the additional RP 5' sequence is listed in SEQ ID No 27.

The synthesis of these primers was performed following the phosphoramidite method, on a GENSET UFPS 24.1 synthesizer.

DNA amplification was performed on a Genius II thermocycler. After heating at 95°C for 10 min, 40 cycles were performed. Each cycle comprised: 30 sec at 95°C, 54°C for 1 min, and 30 sec at 72°C. For final elongation, 10 min at 72°C ended the amplification. The quantities of the amplification products obtained were determined on 96-well microtiter plates, using a fluorometer and Picogreen as intercalant agent (Molecular Probes).

EXAMPLE 4 : IDENTIFICATION OF BIALLELIC MARKERS: SEQUENCING OF AMPLIFIED GENOMIC DNA AND IDENTIFICATION OF POLYMORPHISMS.

The sequencing of the amplified DNA obtained in example 3 was carried out on ABI 377 sequencers. The sequences of the amplification products were determined using automated dideoxy terminator sequencing reactions with a dye terminator cycle sequencing protocol. The products of the sequencing reactions were run on sequencing gels and the sequences were determined using gel image analysis (ABI Prism DNA Sequencing Analysis software (2.1.2 version)).

The sequence data were further evaluated using the above mentioned polymorphism analysis software designed to detect the presence of biallelic markers among the pooled amplified fragments. The polymorphism search was based on the presence of superimposed peaks in the electrophoresis pattern resulting from different bases occurring at the same position as described previously.

11 fragments of amplification were analyzed. In these segments, 13 biallelic markers referred to as A1 to A13 in the BM column were detected. The localization of these biallelic markers is as shown in Table 2.

Table 2

Amplicon	BM	Marker Name	Localization in *OLF* gene cluster	Polymor-phism	BM position in SEQ ID No 1
99-13670	A1	99-13670-305	Between Orf1 and Orf2	A/C	7521
99-13669	A2	99-13669-471	Between Orf1 and Orf2	A/C	8192
99-13666	A3	99-13666-275	Between Orf2 and Orf3	A/T	14483
99-13664	A4	99-13664-221	Between Orf2 and Orf3	A/G	19625
99-13663	A5	99-13663-218	Between Orf2 and Orf3	C/T	20583
99-13660	A6	99-13660-277	Between Orf4 and Orf5	G/T	76947
99-13652	A7	99-13652-407	Between Orf5 and Orf6	G/C	91088
99-13652	A8	99-13652-357	Between Orf5 and Orf6	C/T	91138
99-13652	A9	99-13652-308	Between Orf5 and Orf6	C/T	91187
99-13671	A10	99-13671-396	Between Orf9 and Orf10	C/T	133998
99-13649	A11	99-13649-286	Between Orf9 and Orf10	A/G	140066
99-13648	A12	99-13648-259	Between Orf9 and Orf10	C/T	141176
99-13647	A13	99-13647-278	After Orf10	C/T	144033

Table 3

BM	Marker Name	Position range of probes in SEQ ID No 1		Probes
A1	99-13670-305	7498	7544	P1
A2	99-13669-471	8169	8215	P2
A3	99-13666-275	14460	14506	P3
A4	99-13664-221	19602	19648	P4
A5	99-13663-218	20560	20606	P5
A6	99-13660-277	76924	76970	P6
A7	99-13652-407	91065	91111	P7
A8	99-13652-357	91115	91161	P8
A9	99-13652-308	91164	91210	P9
A10	99-13671-396	133975	134021	P10
A11	99-13649-286	140043	140089	P11
A12	99-13648-259	141153	141199	P12
A13	99-13647-278	144010	144056	P13

EXAMPLE 5 : VALIDATION OF THE POLYMORPHISMS THROUGH MICROSEQUENCING

The biallelic markers identified in example 4 were further confirmed and their respective frequencies were determined through microsequencing. Microsequencing was carried out for each individual DNA sample described in Example 2.

Amplification from genomic DNA of individuals was performed by PCR as described above for the detection of the biallelic markers with the same set of PCR primers (Table 1).

The preferred primers used in microsequencing were about 19 nucleotides in length and hybridized just upstream of the considered polymorphic base. According to the invention, the primers used in microsequencing are detailed in Table 4.

Table 4

Marker Name	BM	Mis. 1	Position range of microsequencing primer mis 1 in SEQ ID No 1		Mis. 2	Complementary position range of microsequencing primer mis. 2 in SEQ ID No 1	
99-13670-305	A1	D1	7502	7520	E1	7522	7540
99-13669-471	A2	D2	8173	8191	E2	8193	8211
99-13666-275	A3	D3	14464	14482	E3	14484	14502
99-13664-221	A4	D4	19606	19624	E4	19626	19644
99-13663-218	A5	D5	20564	20582	E5	20584	20602
99-13660-277	A6	D6	76928	76946	E6	76948	76966
99-13652-407	A7	D7	91069	91087	E7	91089	91107
99-13652-357	A8	D8	91119	91137	E8	91139	91157
99-13652-308	A9	D9	91168	91186	E9	91188	91206
99-13671-396	A10	D10	133979	133997	E10	133999	134017
99-13649-286	A11	D11	140047	140065	E11	140067	140085
99-13648-259	A12	D12	141157	141175	E12	141177	141195
99-13647-278	A13	D13	144014	144032	E13	144034	144052

Mis 1 and Mis 2 respectively refer to microsequencing primers which hybridized with the non-coding strand of the olfactory receptor gene or with the coding strand of the olfactory receptor gene.

The microsequencing reaction was performed as follows :

After purification of the amplification products, the microsequencing reaction mixture was prepared by adding, in a 20µl final volume: 10 pmol microsequencing oligonucleotide, 1 U Thermosequenase (Amersham E79000G), 1.25 µl Thermosequenase buffer (260 mM Tris HCl pH 9.5, 65 mM $MgCl_2$), and the two appropriate fluorescent ddNTPs (Perkin Elmer, Dye Terminator Set 401095) complementary to the nucleotides at the polymorphic site of each biallelic marker tested, following the manufacturer's recommendations. After 4 minutes at 94°C, 20 PCR cycles of 15 sec at 55°C, 5 sec at 72°C, and 10 sec at 94°C were carried out in a Tetrad PTC-225 thermocycler (MJ Research). The unincorporated dye terminators were then removed by ethanol precipitation. Samples were finally resuspended in formamide-EDTA loading buffer and heated for 2 min at 95°C before

being loaded on a polyacrylamide sequencing gel. The data were collected by an ABI PRISM 377 DNA sequencer and processed using the GENESCAN software (Perkin Elmer).

Following gel analysis, data were automatically processed with software that allows the determination of the alleles of biallelic markers present in each amplified fragment.

The software evaluates such factors as whether the intensities of the signals resulting from the above microsequencing procedures are weak, normal, or saturated, or whether the signals are ambiguous. In addition, the software identifies significant peaks (according to shape and height criteria). Among the significant peaks, peaks corresponding to the targeted site are identified based on their position. When two significant peaks are detected for the same position, each sample is categorized classification as homozygous or heterozygous type based on the height ratio.

While the preferred embodiment of the invention has been illustrated and described, it will be appreciated that various changes can be made therein by the one skilled in the art without departing from the spirit and scope of the invention.

REFERENCES :

Adra et al.,1987, Gene, **60**:65-74

Beard et al., 1980, Virology, Vol. 75:81

Beaucage et al., *Tetrahedron Lett* 1981, **22**: 1859-1862

Bolmont et al., J. of Submicroscopic cytology and pathology, 1990, **22**:117-122

Brown EL, Belagaje R, Ryan MJ, Khorana HG, *Methods Enzymol* 1979;**68**:109-151

Chai H. et al., 1993, Biotechnol. Appl. Biochem., **18**:259-273

Cherif D, Julier C, Delattre O, Derré J, Lathrop GM and Berger R, *P.N.A.S. USA* (1990) **87**: 6639-6643

Compton J. *Nature*. 1991 Mar 7; **350**(6313): 91-92.

Current Protocols in Molecular Biology, 1989, Ausubel FM et al. (eds), Greene Publishing Associates, Sections 9.10-9.14.

Davis, L. et al. Basic Methods in Molecular Biology Elsevier, New York. Section 21-2

Engvall E, 1980, Methods Enzymol, **70**(A):419-439

Feldman and Steg, 1996, Medecine/Sciences, synthese, **12**:47-55

Felgner et al., 1987, Proc. Natl. Acad. Sci., **84**:7413

Fisher, D., Chap. 42 in: Manual of Clinical Immunology, 2d Ed. (Rose and Friedman, Eds.) Amer. Soc. For Microbiol., Washington, D.C. (1980).

Flotte et al., 1992, Am. J. Respir. Cell Mol. Biol., **7** : 349-356.

Fraley et al., 1980, J. Biol. Chem., **255**:10431).

Fuller S.A. et al., 1996, Immunology in Current Protocols in Molecular Biology, Ausubel et al. Eds, John Wiley & Sons, Inc., USA

Graham, 1984, EMBO J., 3:2917

Guatelli JC et al., 1990, Proc. Natl. Acad. Sci. USA, 87 : 1874-1878.

Guzman, 1981, Cell, 23 : 175

Hames BD and Higgins SJ, 1985, "Nucleic acid hybridization : a practical approach", Hames and Higgins Ed., IRL Press, Oxford.

Harju L, et al., *Clin Chem* 1993;39(11Pt 1):2282-2287

Hartley JW, Rowe WP, *J Virol* 1976;19(1):19-25

Hillier L. and Green P. *Methods Appl.*, 1991, 1: 124-8.

Houbenweyl, 1974, in Meuthode der Organischen Chemie, E. Wunsch Ed., Volume 15-I et 15-II, Thieme, Stuttgart

Huygen et al., 1996, Nature Medicine, 2(8):893-898

Julan et al., 1992, J. Gen. Virol., 73 : 3251 – 3255.

Kaneda et al., 1989, Science, 243:375

Kiefer H et al., 1996, Biochemistry, 35(50):16077-16084.

Koch Y., 1977, Biochem. Biophys. Res. Commun., 74:488-491

Kohler G. and Milstein C., 1975, Nature, 256 : 495.

Kort et al., 1983, Nucleic Acids Research, 11:8287-8301

Lenhard T. et al., 1996, Gene, 169:187-190

Levrero et al., 1991, Gene, 101:195

Lin Z, Floros J, 1998, Biotechniques, 24(6):937-940

Livak KJ, and Hainer JW., 1994, Hum Mutat., 3(4): 379-385.

Mackey K, Steinkamp A, Chomczynski P, 1998, Mol Biotechnol, 9(1):1-5

Mansour SL, Thomas KR, Capecchi MR, *Nature* 1988;336(6197):348-52

McLaughlin et al., 1989, J. Virol., 62 : 1963 – 1973.

Merrifield RB 1965 *Nature*. 207: 522-523.

Merrifield RB 1965 *Science*. 150: 178-185.

Midoux, 1993, Nucleic Acids Research, 21:871-878

Muzyczka et al., 1992, Curr. Topics in Micro. and Immunol., 158 : 97-129.

Narang SA, Hsiung HM, Brousseau R, *Methods Enzymol* 1979;68:90-98

Neda et al., 1991, J. Biol. Chem., 266 : 14143 – 14146.

Nekrasova E, Sosinskaya A, Natochin M, Lancet D, Gat U, *Eur J Biochem* 1996;238(1):28-37

O'Reilly et al., 1992, Baculovirus expression vectors : a Laboratory Manual. W.H. Freeman and Co., New York Adra et al., 1987, Gene, 60:65-74

Ohno et al., 1994, Sciences, 265:781-784

Ohwada A, Hirschowitz EA, Crystal RG, *Hum Gene Ther* 1996;7:1567-76

Ouchterlony, O. et al., Chap. 19 in: Handbook of Experimental Immunology D. Wier (ed) Blackwell (1973)

Pagano et al., 1967, J. Virol., 1:891

Palmer S et al., 1989, Biochim Biophys Acta, 1014(3):239-246.

Pastore, 1994, Circulation, 90:I-517

PCR Methods and Applications" (1991, Cold Spring Harbor Laboratory Press

Pearson WR, Lipman DJ, *Proc Natl Acad Sci U S A* 1988;85(8):2444-8

Porath J et al., 1975, Nature, 258(5536) : 598-599.

Raming K et al., 1993, Nature, 361(6410):353-356

Rettlez and Basenga, 1987, Mol. Cell. Biol., 7:1676-1685

Romana SP, Tachdjian G, Druart L, Cohen D, Berger R, Cherif D *Eur J Hum Genet* (1993) 1: 245-251.

Roth J.A. et al., 1996, Nature Medicine, 2(9):985-991

Rougeot, C. et al.,. *Eur. J. Biochem.* 219 (3): 765-773, 1994

Sambrook, J. Fritsch, E. F., and T. Maniatis. 1989. Molecular cloning: a laboratory manual. 2ed. Cold Spring Harbor Laboratory, Cold spring Harbor, New York

Samulski et al., 1989, J. Virol., 63 : 3822-3828.

Sanchez-Pescador R., 1988, J. Clin. Microbiol., 26(10):1934-1938

Sczakiel G. et al., 1995, Trends Microbiol., 1995, 3(6):213-217

Smith et al., 1983, Mol. Cell. Biol., 3:2156-2165.

Steiner AL et al., 1972, J Biol Chem, 247(4):1106-1113

Syvanen AC, et al., 1994, Hum Mutat., 3(3): 172-179.

Tacson et al., 1996, Nature Medicine, 2(8):888-892

Tucker J, Grisshammer R, *Biochem J* 1996;317(Pt 3):891-899

Urdea M.S., 1988, Nucleic Acids Research, 11: 4937-4957

Urdea MS et al., 1991, Nucleic Acids Symp Ser., 24: 197-200.

Vaitukaitis J. et al., 1971, J. Clin. Endocrinol. Metab., 33 : 988-991

Vlasak R. et al., 1983, Eur. J. Biochem., 135:123-126

White, B.A. Molecular Cloning to Genetic Engineering Ed. in Methods in Molecular Biology 67: Humana Press, Totowa 1997

Zhao H et al., 1998, Science, 279(5348):237-242

Zhenlin et al., Gene, 1989, 78:243-254.

SEQUENCE LISTING FREE TEXT

The following free text appears in the accompanying Sequence Listing :

open reading frame

ubiquitin 1 pseudogene complement

ubiquitin 2 pseudogene complement

polymorphic base

or

complement

probe

sequencing oligonucleotide PrimerPU

sequencing oligonucleotide PrimerRP

What is claimed:

1. An isolated, purified, or recombinant polynucleotide comprising a contiguous span of at least 12 nucleotides of SEQ ID No 1 or the complements thereof, wherein said contiguous span comprises at least 1 of the following nucleotide positions of SEQ ID No 1: 1-113643, 114064-127488, 127855-144460.

2. An isolated, purified, or recombinant polynucleotide comprising a contiguous span of at least 12 nucleotides of a sequence selected from the group consisting of SEQ ID Nos 2-11 or the complements thereof.

3. An isolated, purified, or recombinant polynucleotide consisting essentially of a contiguous span of 8 to 50 nucleotides of SEQ ID No 1 or the complement thereof, wherein said span includes an olfactory receptor-related biallelic marker in said sequence.

4. A polynucleotide according to claim 3, wherein said olfactory receptor-related biallelic marker is selected from the group consisting of A1 to A13, and the complements thereof.

5. A polynucleotide according to claims 3 or 4, wherein said contiguous span is 18 to 47 nucleotides in length and said biallelic marker is within 4 nucleotides of the center of said polynucleotide.

6. A polynucleotide according to claim 5, wherein said polynucleotide consists essentially of a sequence selected from the following sequences: P1 to P13, and the complementary sequences thereto.

7. A polynucleotide according to any one of claims 1, 2 or 3, wherein the 3' end of said contiguous span is present at the 3' end of said polynucleotide.

8. A polynucleotide according to claims 3 or 4, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide and said biallelic marker is present at the 3' end of said polynucleotide.

9. An isolated, purified, or recombinant polynucleotide consisting essentially of a contiguous span of 8 to 50 nucleotides of SEQ ID No 1 or the complement thereof, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide, and wherein the 3' end of said

polynucleotide is located within 20 nucleotides upstream of an olfactory receptor-related biallelic marker in said sequence.

10. A polynucleotide according to claim 9, wherein the 3' end of said polynucleotide is located 1 nucleotide upstream of said olfactory receptor-related biallelic marker in said sequence.

11. A polynucleotide according to claim 10, wherein said polynucleotide consists essentially of a sequence selected from the following sequences: D1 to D13, and E1 to E13.

12. A polynucleotide according to claim 7 consisting essentially of a sequence selected from the following sequences: B1 to B11 and C1 to C11.

13. An isolated, purified, or recombinant polynucleotide which encodes a polypeptide comprising a contiguous span of at least 6 amino acids of a sequence selected from the group consisting of SEQ ID Nos 12-21.

14. A polynucleotide for use in a genotyping assay for determining the identity of the nucleotide at an olfactory receptor-related biallelic marker or the complement thereof.

15. A polynucleotide according to claim 14, wherein the polynucleotide is used in an assay selected from the group consisting of: a hybridization assay, a sequencing assay, an enzyme-based mismatch detection assay, and an amplification of a segment of nucleotides comprising said biallelic marker.

16. A polynucleotide according to any one of claims 1-15 attached to a solid support.

17. An array of polynucleotides comprising at least one polynucleotide according to claim 16.

18. An array according to claim 17, wherein said array is addressable.

19. A polynucleotide according to any one of claims 1-15, further comprising a label.

20. A recombinant vector comprising a polynucleotide according to any one of claims 1-15.

21. A host cell comprising a recombinant vector according to claim 20.

22. A non-human host animal or mammal comprising a recombinant vector according to claim 20.

23. A mammalian host cell comprising an olfactory receptor gene disrupted by homologous recombination with a knock out vector, comprising a polynucleotide according to any one of claims 1-15.

24. A non-human host mammal comprising an olfactory receptor gene disrupted by homologous recombination with a knock out vector, comprising a polynucleotide according to any one of claims 1-15.

25. An isolated, purified, or recombinant polypeptide comprising a contiguous span of at least 6 amino acids of a sequence selected from the group consisting of SEQ ID Nos 12-21.

26. An isolated or purified antibody composition are capable of selectively binding to an epitope-containing fragment of a polypeptide according to claim 25.

27. A method of genotyping comprising determining the identity of a nucleotide at an olfactory receptor-related biallelic marker or the complement thereof in a biological sample.

28. A method according to claim 27, wherein said biological sample is derived from a single subject.

29. A method according to claim 28, wherein the identity of the nucleotides at said biallelic marker is determined for both copies of said biallelic marker present in said individual's genome.

30. A method according to claim 27, wherein said biological sample is derived from multiple subjects.

31. A method according to claim 27, further comprising amplifying a portion of said sequence comprising the biallelic marker prior to said determining step.

32. A method according to claim 31, wherein said amplifying step is performed by PCR.

33. A method according to claim 27, wherein said determining is performed by an assay selected from the group consisting of: a hybridization assay, a sequencing assay, a microsequencing assay, and an enzyme-based mismatch detection assay.

34. A method according to claim 27 wherein said olfactory receptor-related biallelic marker is selected from the group consisting of A1 to A13 and the complements thereof.

35. A method for the screening of a candidate substance interacting with an olfactory receptor polypeptide selected from the group consisting of SEQ ID Nos 12-21, or fragments or variants thereof, comprises the following steps :

a) providing a polypeptide selected from the group consisting of the sequences of SEQ ID Nos 12-21 , or a peptide fragment or a variant thereof;

b) obtaining a candidate substance;

c) bringing into contact said polypeptide with said candidate substance; and

d) detecting the complexes formed between said polypeptide and said candidate substance.

36. A method for the screening of ligand molecules interacting with an olfactory receptor polypeptide selected from the group consisting of SEQ ID Nos 12-21, wherein said method comprises :

a) providing a recombinant eukaryotic host cell containing a nucleic acid encoding a polypeptide selected from the group consisting of the polypeptides comprising the amino acid sequences SEQ ID Nos 12-21;

b) preparing membrane extracts of said recombinant eukaryotic host cell;

c) bringing into contact the membrane extracts prepared at step b) with a selected ligand molecule; and

d) detecting the production level of second messengers metabolites.

37. A method for the screening of ligand molecules interacting with an olfactory receptor polypeptide selected from the group consisting of SEQ ID Nos 12-21, wherein said method comprises :

a) providing an adenovirus containing a nucleic acid encoding a polypeptide selected from the group consisting of the polypeptides comprising the amino acid sequences SEQ ID Nos 12-21;

b) infecting an olfactory epithelium with said adenovirus;

c) bringing into contact the olfactory epithelium b) with a selected ligand molecule; and

d) detecting the increase of the response to said ligand molecule.

FIGURE 1

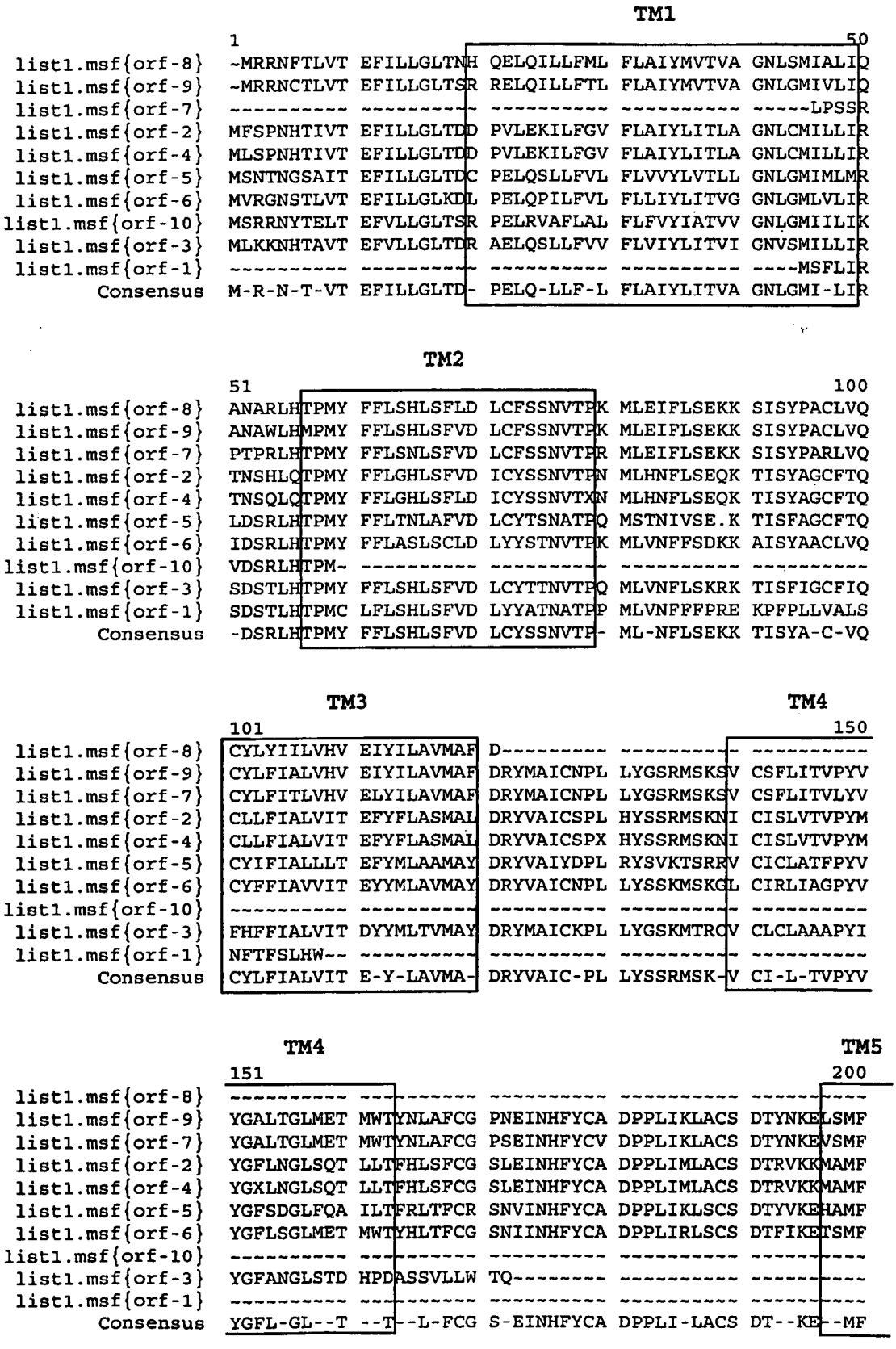
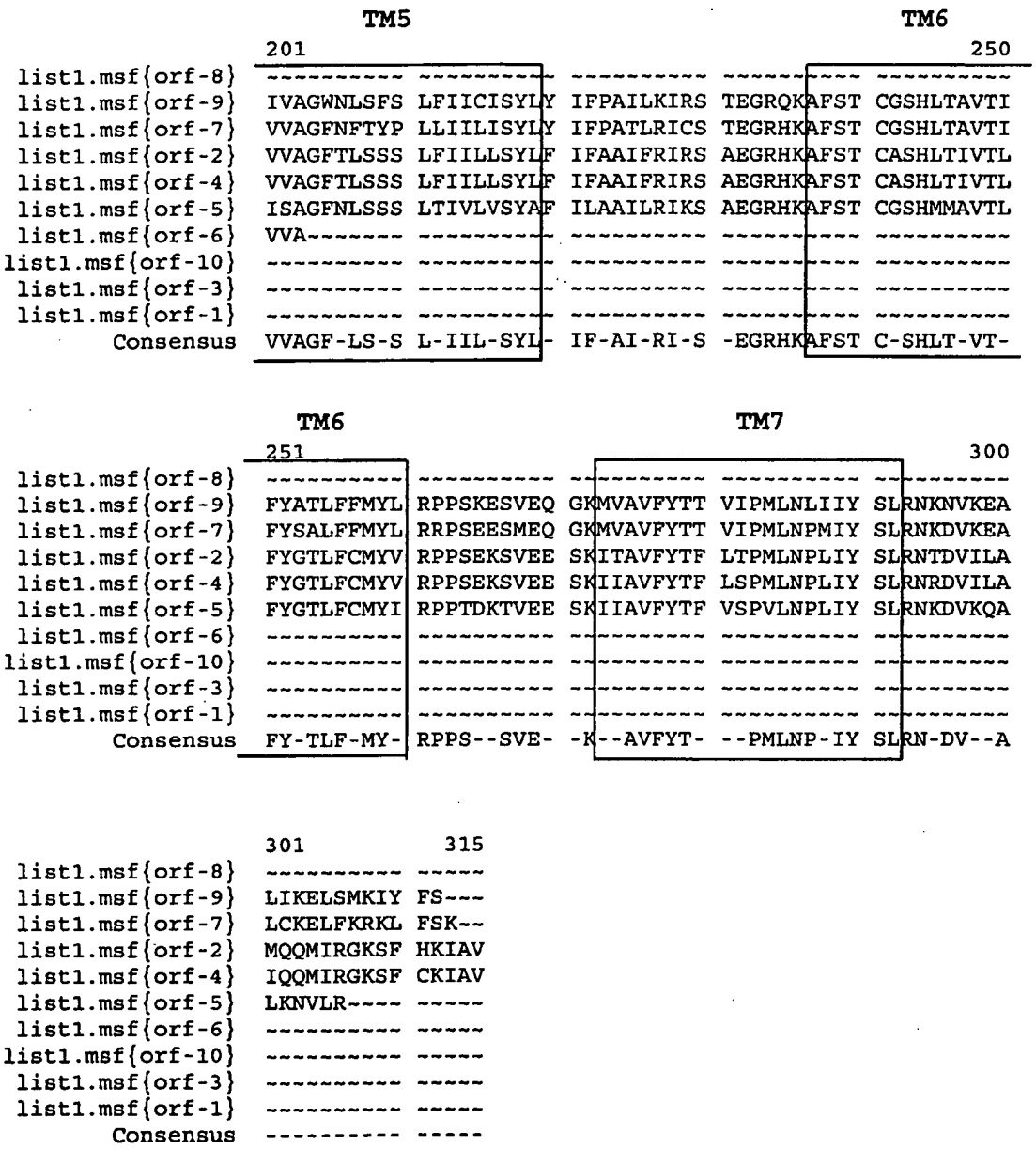

TM1
TM2
TM3
TM4
TM4
TM5
list1.msf{orf-8}
list1.msf{orf-9}
list1.msf{orf-7}
list1.msf{orf-2}
list1.msf{orf-4}
list1.msf{orf-5}
list1.msf{orf-6}
list1.msf{orf-10}
list1.msf{orf-3}
list1.msf{orf-1}
Consensus

FIGURE 1 (continued)



<110> Genset SA

<120> Genes encoding olfactory receptors and biallelic markers thereof.

<150> US 60/104,299
<151> 1999-10-13

<160> 27

<170> Patent.pm

<210> 1
<211> 144460
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 2406..2600
<223> open reading frame 1

<220>
<221> CDS
<222> 9711..10658
<223> open reading frame 2

<220>
<221> CDS
<222> 24851..25369
<223> open reading frame 3

<220>
<221> CDS
<222> 45714..46661
<223> open reading frame 4

<220>
<221> CDS
<222> 80198..81115
<223> open reading frame 5

<220>
<221> CDS
<222> 96291..96902
<223> open reading frame 6

<220>
<221> CDS
<222> 110758..111564
<223> open reading frame 7

<220>
<221> CDS
<222> 122525..122887
<223> open reading frame 8

<220>
<221> CDS
<222> 132454..133389
<223> open reading frame 9

<220>

<221> CDS
<222> 143398..143577
<223> open reading frame 10

<220>
<221> misc_feature
<222> 113644..114063
<223> ubiquitin 1 pseudogene complement

<220>
<221> misc_feature
<222> 127489..127854
<223> ubiquitin 2 pseudogene complement

<220>
<221> allele
<222> 7521
<223> 99-13670-305 : polymorphic base G or T

<220>
<221> allele
<222> 8192
<223> 99-13669-471 : polymorphic base G or T

<220>
<221> allele
<222> 14483
<223> 99-13666-275 : polymorphic base A or T

<220>
<221> allele
<222> 19625
<223> 99-13664-221 : polymorphic base C or T

<220>
<221> allele
<222> 20583
<223> 99-13663-218 : polymorphic base A or G

<220>
<221> allele
<222> 76947
<223> 99-13660-277 : polymorphic base A or C

<220>
<221> allele
<222> 91088
<223> 99-13652-407 : polymorphic base G or C

<220>
<221> allele
<222> 91138
<223> 99-13652-357 : polymorphic base A or G

<220>
<221> allele
<222> 91187
<223> 99-13652-308 : polymorphic base A or G

<220>
<221> allele
<222> 133998
<223> 99-13671-396 : polymorphic base A or G

<220>
<221> allele
<222> 140066
<223> 99-13649-286 : polymorphic base C or T

<220>
<221> allele
<222> 141176
<223> 99-13648-259 : polymorphic base A or G

<220>
<221> allele
<222> 144033
<223> 99-13647-278 : polymorphic base A or G

<220>
<221> primer_bind
<222> 7362..7380
<223> 99-13670.rp

<220>
<221> primer_bind
<222> 7805..7824
<223> 99-13670.pu complement

<220>
<221> primer_bind
<222> 8120..8140
<223> 99-13669.rp

<220>
<221> primer_bind
<222> 8643..8662
<223> 99-13669.pu complement

<220>
<221> primer_bind
<222> 14308..14328
<223> 99-13666.rp

<220>
<221> primer_bind
<222> 14740..14757
<223> 99-13666.pu complement

<220>
<221> primer_bind
<222> 19346..19366
<223> 99-13664.rp

<220>
<221> primer_bind
<222> 19826..19845
<223> 99-13664.pu complement

<220>
<221> primer_bind
<222> 20298..20318
<223> 99-13663.rp

<220>
<221> primer_bind

<222> 20781..20800
<223> 99-13663.pu complement

<220>
<221> primer_bind
<222> 76752..76772
<223> 99-13660.rp

<220>
<221> primer_bind
<222> 77205..77223
<223> 99-13660.pu complement

<220>
<221> primer_bind
<222> 90967..90987
<223> 99-13652.rp

<220>
<221> primer_bind
<222> 91474..91494
<223> 99-13652.pu complement

<220>
<221> primer_bind
<222> 133925..133945
<223> 99-13671.rp

<220>
<221> primer_bind
<222> 134375..134393
<223> 99-13671.pu complement

<220>
<221> primer_bind
<222> 139807..139826
<223> 99-13649.rp

<220>
<221> primer_bind
<222> 140331..140351
<223> 99-13649.pu complement

<220>
<221> primer_bind
<222> 140912..140932
<223> 99-13648.rp

<220>
<221> primer_bind
<222> 141416..141434
<223> 99-13648.pu complement

<220>
<221> primer_bind
<222> 143828..143847
<223> 99-13647.rp

<220>
<221> primer_bind
<222> 144292..144309
<223> 99-13647.pu complement

<220>
<221> misc_binding
<222> 7498..7544
<223> 99-13670-305.probe

<220>
<221> misc_binding
<222> 8169..8215
<223> 99-13669-471.probe

<220>
<221> misc_binding
<222> 14460..14506
<223> 99-13666-275.probe

<220>
<221> misc_binding
<222> 19602..19648
<223> 99-13664-221.probe

<220>
<221> misc_binding
<222> 20560..20606
<223> 99-13663-218.probe

<220>
<221> misc_binding
<222> 76924..76970
<223> 99-13660-277.probe

<220>
<221> misc_binding
<222> 91065..91111
<223> 99-13652-407.probe

<220>
<221> misc_binding
<222> 91115..91161
<223> 99-13652-357.probe

<220>
<221> misc_binding
<222> 91164..91210
<223> 99-13652-308.probe

<220>
<221> misc_binding
<222> 133975..134021
<223> 99-13671-396.probe

<220>
<221> misc_binding
<222> 140043..140089
<223> 99-13649-286.probe

<220>
<221> misc_binding
<222> 141153..141199
<223> 99-13648-259.probe

<220>
<221> misc_binding
<222> 144010..144056

<223> 99-13647-278.probe

<220>
<221> primer_bind
<222> 7502..7520
<223> 99-13670-305.mis

<220>
<221> primer_bind
<222> 7522..7540
<223> 99-13670-305.mis complement

<220>
<221> primer_bind
<222> 8173..8191
<223> 99-13669-471.mis

<220>
<221> primer_bind
<222> 8193..8211
<223> 99-13669-471.mis complement

<220>
<221> primer_bind
<222> 14464..14482
<223> 99-13666-275.mis

<220>
<221> primer_bind
<222> 14484..14502
<223> 99-13666-275.mis complement

<220>
<221> primer_bind
<222> 19606..19624
<223> 99-13664-221.mis

<220>
<221> primer_bind
<222> 19626..19644
<223> 99-13664-221.mis complement

<220>
<221> primer_bind
<222> 20564..20582
<223> 99-13663-218.mis

<220>
<221> primer_bind
<222> 20584..20602
<223> 99-13663-218.mis complement

<220>
<221> primer_bind
<222> 76928..76946
<223> 99-13660-277.mis

<220>
<221> primer_bind
<222> 76948..76966
<223> 99-13660-277.mis complement

<220>

<221> primer_bind
<222> 91069..91087
<223> 99-13652-407.mis

<220>
<221> primer_bind
<222> 91089..91107
<223> 99-13652-407.mis complement

<220>
<221> primer_bind
<222> 91119..91137
<223> 99-13652-357.mis

<220>
<221> primer_bind
<222> 91139..91157
<223> 99-13652-357.mis complement

<220>
<221> primer_bind
<222> 91168..91186
<223> 99-13652-308.mis

<220>
<221> primer_bind
<222> 91188..91206
<223> 99-13652-308.mis complement

<220>
<221> primer_bind
<222> 133979..133997
<223> 99-13671-396.mis

<220>
<221> primer_bind
<222> 133999..134017
<223> 99-13671-396.mis complement

<220>
<221> primer_bind
<222> 140047..140065
<223> 99-13649-286.mis

<220>
<221> primer_bind
<222> 140067..140085
<223> 99-13649-286.mis complement

<220>
<221> primer_bind
<222> 141157..141175
<223> 99-13648-259.mis

<220>
<221> primer_bind
<222> 141177..141195
<223> 99-13648-259.mis complement

<220>
<221> primer_bind
<222> 144014..144032
<223> 99-13647-278.mis

<220>
<221> primer_bind
<222> 144034..144052
<223> 99-13647-278.mis complement

<400> 1
caattaaagt tttgttcact ataagtcttt tttggaaaag agagagaaac attcaaatta 60
tttacatacc agtttccatt agcatgtgaa gaacaaacag aaacactttt cagggtgaac 120
aaaattcctg ctacagttat aaaatcctgc atatactctt tactttgtga ttctgaaaaa 180
caccgttcta cctggtttat tgaaatgtgt gaaagctcta atgcaatgtt atttttaca 240
ttttgtaaca cttaagtcat aaagccaagc tattctcaaa ccttgatgaa acatgttgga 300
agaaattatg ttttagtgtt tggtgaaaac attatgtttc gtcacttaag gtgataaatt 360
gtactcatta aagaactttg aaagttcaca catagccaat ggtttaaaat gcactaattt 420
agattcccaa ttctcacaaa ggccagttac tctggaccat tcaatcgcca aagaggaaaa 480
ctgggggcat tcccatcccg ggatatggga agtccccgag cttccagcct ggtccttgtg 540
gccgaaaaat ggcatgtttt gcttttgctt ttggatcctg ttcgtgccgc caaaaatgtt 600
ctgtgtgggg aaagtgcgag gggagagaaa agacacggac atgatacgtt taagggtaaa 660
caacgtttat cccatgtaag tggccatgca gatatagtaa gcaaatgata taataataag 720
caaatgatat aataagcaga ttgatataat aagtagattg caatggaacg gggaaaaggg 780
aaaatacatc tacattcacc agactatgga ggattcaaca acagactggg acgcaacagc 840
ctgggctcca gagtcagata ggtaggcaaa gagatcctag ttctatacag atacgtacca 900
tggagcagtt ccactttcct aagcacattc agttgtgata aaaatagatg agtttcaagg 960
gctgatacat tacatgccac actcaaagtt gtgttgttaa acaatttcaa ttgttgttac 1020
aatttcaaat aaaagcaatg tttacaacca tgggttcaag agaagtctaa gtgaacacat 1080
ataataaaga cttgcaaaat aataaaagat aaggctcttt aactatcaaa agacttgcag 1140
aaaagaacca cagaaaacca ttttaaatat aactgccttc gtatgtaaga aattctacat 1200
tatttttgat gttaaaacat caatctcatg cttactaggc tatttcttaa tgacacatgt 1260
atttacaaat ttgagagaag aggaagaaat atcaggtgac accactgggt taatgcataa 1320
atgacaaacc taaatgcatt ttaatttcct tttctttaaa tcgagctgag cttcagcccc 1380
ttctttttgt ggtgttctta gtcatctacc ttatcacagt aatcacaatg taagcatgat 1440
cttctttttt ttttttaagt gcacaatatt tttaactgtt aacaatatac ctattgttac 1500
ctatgggcac aatgatatac agcatatctc tagaatttat tcttgcaaaa ctataacttt 1560
atacctgctg aacagcaaca ccccatttct ccctttcctc cagccgctgc aaccaccttc 1620
tattctctgt ttctatgagt ttgactattt tggattcctc atataaattt aatcatgcag 1680
tatttgtcct tccgtgcctg gcttatttca cttaacataa tgtcctccag gttcatcata 1740
tgacaggatt tcttcttttt cttaatgatg aataatattc cattacatgt gtgtactaca 1800
ttttcttcat ctttcaatgg acatttaggt tgtttctata tctggactat tgtaaataat 1860
ggtgcaatga acataagagt acctatgtct cttcaagagc ttgatttaaa ttcttttgga 1920
tatatgccca gaagtgcaat tgctggttta tatgataatt cgatttttaa ttatttgaag 1980
actcatcata ctgtttttta tagtggctgc acaattttat attcccacca atgttgtaca 2040
agggttccaa tttcttcata tgtcaccaat atttgttgtc ttttggattt ttttaaaata 2100
aagtaacagc catcataaca aatgtgatat catgcttttg tttcatatgc attttcctga 2160
tgattagtgt gttgagcacc ttttcatttt tatttattta tttatttata ctctaagttc 2220
tgggatacat ctgcagaaca cgcaggtttg ttacataggt atacatgtgc catggtggtt 2280
tgctgcgccc attaacctgt catctacatt aggtatttct gctaatacta tccctccccc 2340
agcccccgac cccctgacag gtcctggtgt gtgatgtttc ccttcctgtg tccatgtgtg 2400
tgagcatgag cttcttaata agaagtgatt caacactaca cactccaatg tgcttgttcc 2460
tcagtcatct ctcctttgta gatctctatt atgccaccaa tgccactcct ccgatgctgg 2520
ttaacttttt ttttccaaga gaaaaaccgt ttcctttatt ggttgcttta tccaatttca 2580
ccttttcatt gcactggtga tcacagatta tcatatgctc acagtgatgg tgtatgacca 2640
ctacatggcc atctgcaagc ctttgttata tggaagcaaa atgtccaggt gtgtctgcct 2700
ctgtctcact gctgctccct atatttatgg ctctgcaaat ggtctggtac aggtcatcct 2760
gatgctttgt ctgttcttct gtgaacccaa tgagatcaac cacttttttt ttttttggag 2820
aaaatgcatt atatgcacat ttaattccac tataaatttt tgaatggacg gttggagagg 2880
aagggagaaa tacatattaa cggagagaat accacccaga aagtatatac aatgggagaa 2940
aggaacctgt tgatccaagt ttccatattc ttattatggc atataaggtc atgattattt 3000
tctcagtatg aagcatctcc cagggctgac tctgatgtaa aattggagat caaccacttt 3060
tattatgcag aaccacccct cttagtcctc gcctgcttgg atacttatgt caaagaaact 3120
gccatgttca tggtggctgg ttccaacctc atctgccctc tcactatcat ctttatttcc 3180
tacactttca tcttcacaga cattctgcat atctgcactg ctgagggaag gtacaatgcc 3240
ttctccacct gcgggtccct tgtgactgcc gtcactgtct ttcaaggaac gctgtttcac 3300
atgtgcctga ggcccccttc tgaggcatct gtagaacagg ggaaaattgt agctgctttt 3360

tatatctttg tgagtcctac gttaaaccca ttgatctacc gtctgaggaa taaaaatgtt 3420
aaaagaacaa taagggaagt tatccaaaag aaactgtttg ctaagtaagg tagatatttt 3480
agttgcaggt tatgtaatac attatttta tcttaccaat taacgagcat tataaattaa 3540
caaatcactt tctgtcattg agtgtttttt gtcttttgta acttgcatat gggaattgaa 3600
agtgtatacc aaattattag ctagagttga cagtgtcatc tcagtgaatt taagaagaaa 3660
tcatagaaat ttaaatagaa gacttatggc atgtaaaagt caataaagaa cagtgattcc 3720
ttctttagta ctcatattgg tagcaaacga taaaagacag aatgcaatgg aaattacagt 3780
tcattacatt tttatagtac ttaataactt ccaaactatt ttctagacac ctttcaaaca 3840
tagtatatga agttttctcc tttcttttat acagataatg caacaataaa gatcactgat 3900
gtagggaaaa gagagatcag actgttactg tgtctatgta gaaaaggaag gcataagaaa 3960
cttcattttg acttgtaccc tgaacaattg ttttgtcctg agatgctgtt aatctgtaac 4020
tttgccccaa ccttgagctt ataaaaacat gtgttgtatg gaatcaaggt ttaagggatc 4080
tagggctgtg caggatgtgc cttgttagca gaatgtatac aggcagtatg cttggtaaaa 4140
gtcatcgcca ttctccagtc tccataaacc aggggcacaa tgcactgtgg aaagtcacag 4200
ggacctctgc cctggaaagc cgggtattgc caaggtttct ccccatgtga tagtctgaaa 4260
tatggcctcg tgggatggga aagacctgac cggcccccag cctgacaccc gtgaagggtc 4320
tgtgctcggg aggattagta aaagaggatg gcctcttata gctgagataa gaggaaggcc 4380
tctgtctcct gcctgcccct gggaactgaa tgtctcggta taaaacctga ttgcaccttt 4440
cttctattct gagataggag aaaaaccgcc ctgtggcggg aggcgagaca tgttggcagc 4500
aatgctgctt tgttattctt tactccactg agatgtttgg gtggagagaa ggaaaaatct 4560
ggcttacgtg cacatccagg catagtacct ccccttgaac ttatttgtga cacagattcc 4620
tttgctcaca tgttttcttg ctgaccttct ccctattatc accctgttct cctaccacat 4680
tcctcttgct gagatagtga aaatagtaat caataaaaac tgagggatct cagagaccgg 4740
tgccgatgca ggtcttccat atgctgagcg ccagtcccct ggggccactg ttctttctct 4800
atactttgtc tctgtgtctt atttcttttc tcagtctctc atcccacctg acgagatata 4860
cccacaggtg cagaggggca ggccacccct tcaattgaag tatatctcag aatactactt 4920
tgagatacag tcttagaatt atattttgag ccaatgaaat cttctttctt gaagcttttg 4980
aagcaatgcc aaatttccgt tagtaggctt tataaatatc attgtttgca ttaccaggag 5040
gcattcacat caatatgtga cctcacttct ccactctttc attgccattg aagcagatac 5100
tttcaagtat gtcttaatat attgattttt atcttctcat tgggggaaca tgggaagtgt 5160
cacatgtggg actacaccgt aatttgggta tttgtagtct taaggttttc atgaagcttc 5220
gtgtgggcct ccatttctct agaacgattt gatgtgttcg ttttttatcc ttcacagcaa 5280
cacatgctta ggcagatgaa tcactgcagc agcatttaga cacatttgtg attcagggat 5340
agatagctct tcagtaggat ggtgtgaatt ttgggataat ggcacatact taaaacagaa 5400
ctacctttg acccagcaat cccattactg ggtatataca ccaaggaata taaatcattc 5460
caccataaag acacatgcat gtgtatgttc atcacaacag tattcacaat agcaaagaca 5520
tggaatcaac ctaaatgccc ttcaacagta gattggatac aaaaaaatgt ggtatatata 5580
catcatggaa tgctatgcag ccgtaaaaaa aaagaatgag attatgtctt ttgtagcaac 5640
acggatgaag ctggaggaca atatccaaag caaaccaatg caggaacagg aaaccaaatg 5700
ctgcaagttc tcacttacat gtggaagcta aacattgaat acacatggac acaaaaaagg 5760
gaacaacagg caccaggacc tacttgaggg tgaagtgtgg gaggagggta aggattgaaa 5820
ctctgcctat caggcactgt gcttatcaac tgggtgatga ataaacctgt acaccaaaac 5880
cctgtgacat ggaatttacc tttataacaa agctgcacat ggacccctga acctaaaata 5940
aaagttaaaa aaagaaatct gtcccaagga gactgttttc tcttaatgtg ctgcatcctg 6000
cttaatgaac tatggatttc atgcattctt tttcaagatt atattgccta cctgattgta 6060
gacagtttga tgcattttac atagtatcag ttaaacatta aacataatta agagcatttg 6120
gcttcaaaat aatagtaaat gggtagaatt tattatggtt atagtactac tcatacaaat 6180
aataatacaa catcagtgat gtagtgtcta gtgagcatga cactattata gaacacttct 6240
taggctggat tttgataata atagcatgct ataacttttg aataaaaata gtaaattgaa 6300
taatcacaaa caagtaaaaa tctaaagggc cagtagtatg tattcaacta gcttaccagt 6360
gtatcatttg tgtagctaaa tccattcgct gtccctctag cagacacaca tgctagttat 6420
tgcagtggaa gaaaaatgaa atgaactaag gaattaatgt ctttgagtaa tataaacaga 6480
gccttctgga ggttttctat gaaaaataac ataagtatgt gtaaaactct tctttgagta 6540
agtatcatac acatagctaa attctgtatt ttctattcat tgctgtataa ataaattaac 6600
attgccttct ttcgtgtgga catctgatca tgtgattcca tacttaaaat tacatatact 6660
ttgctctttc taaatgaaaa tagcccattc atctatactt ttcatctaac cattcagtta 6720
ttatagggct agttttattt ttgtcagaaa attcactttg taaatcttat gttttattat 6780
gtagccattg tataccaaca tataaaagaa aatacataca tacatacaca catacattag 6840
ggaatttttt cttcaaagtg aaagcagtct acttaaaaat tatccaaatc ttcaacattt 6900
ttatgcaaaa caaggtacct ctacattaag ggaggaggaa atgagggaca ccagttttat 6960
aatcttatga taccttatat tccatcttaa ttttttttt tgagatgaag tctcgctctg 7020
tcacccaggc tgaggtgcag tggtgtaatt ttggctcatt gcaacctctg cctcctgggt 7080
tcaagcaatt ctccctgcct cagcctcctg agtagctggg attacaggca cgcaccacca 7140

aacctggcta agttttgtat ttttagtaga gacagggttt cgccatgttg gccaggctgg 7200
cctcaaactc ttgacctcag gtgatctagc ggccttggcc tcccaaagtg ctgggattat 7260
aggcaagagc caccgtgccc ggcttccaaa atatttaagc aatattattg cattttacaa 7320
ttttagtaat gcaaggaacc aaaataaaac agaataattg agatagagaa gactttacac 7380
atcaatttga aataaacata taggaatgcc ctatattctc aaatttcatg ggatgtaaca 7440
aatctacatt ttgattttga tatatagcta tattttatta aatgatttct cagagtataa 7500
cccaccaccc gcaactctaa kaaaattagg gatgattctc cgtcttggtc agactgtact 7560
ttgatccatt tgtgctaacc tggaactata tgtgcactgg aagatacaga ctaatgaacg 7620
catctctagg tccctttgtc ctccaacaaa tacagtgcta cacaattttc aaatattctc 7680
attctatttg caattccctt tctaaatcaa caatttatg catcatcaat tttataaata 7740
gccctggttc tgagaccttt gatgattagc atgttaatat taaccttgat agtgcagact 7800
agttccaaaa taaatccata acgcccgtcc tccaaaggat cctgggccct gaccaatact 7860
tctgcctcct tctactgatt atgccactct tctgctctca cagtctttga aatctttcat 7920
cttctcaaat gtctcacctg ccacaacttc ctacccctca ctcattagat gacctcactg 7980
cagtgtttat aggcaaaagc tgaagagttc tgatgagaat tccttgtttt catcatacta 8040
aaaaggcaga agttatctac ccatgttttc ctcttccatg ttattgaaag ggatggctaa 8100
tttctctgtt tgtactaagg atctcatgca cacttgtacc gaaactctac ataaatgtaa 8160
aataggaatt ttatcctcgc aggagagtct akctccagga agcatcaaat gaaggttatg 8220
ctcttgagga acaaagcata ttttaattgg ctctttagaa ttagtttaaa aatactctag 8280
ggaagatacc taataaatat attgtaaatc ttgctacctt gttttcataa gatatggtcc 8340
attaacatta acaggtaggt cattttctac atttgttcac taaaaataca atttgtgtgt 8400
gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtagag ataagtattc catcaatggg 8460
gaaaacaaat attttctcca atttcactag acttttctat ccatatgtta gtattgcatt 8520
gactctctac aatcaattat atgtaagctc ttttgatgta tacatgagaa ctaaataaac 8580
aatggcttta aggtggaatt tttggagtag aaatttttac aatcttatgt tactttcaaa 8640
gagaactgag agggagtggc ttgagttgga agctaagacc ccttttaaaa attatctgtg 8700
catacctaaa caatattgtc atgagtgggt tagtggcaca aaaatgggaa caaaaaaaat 8760
ccacactttc cccaaatagt agaattgcat atctgctgtt taaatgctgc attcctagaa 8820
agatcttctt tgagtggtca tctaaaagca tcctacattc atgttttctt cacagaaaag 8880
tggtttgcct gaaattgtat gtaacttttt ctgtgtgttt aatctctgtt ccctctatta 8940
taatgtgtgc tctgctccct ctattataat gtgtgctccc tgagctggga gacttgttta 9000
ctgtagtaac cgcaggaatt ggaacagaaa gaaatgaagc tatacagtat acatgagtaa 9060
acaggcagtg acattacaaa gtggaaaaaa acaagtcttc attttgtacc ctcttagcca 9120
tatatcagat agaaattaat ttctctagtt taatcgttcc tgaataaagg taaggcacac 9180
aactatgggt cttaattgaa aatgctttgc ttttctttct tcattttgta tctgaaacaa 9240
tacaatatca gagctggagg tataataaag atcagacttc ctttatttat ccatttgaaa 9300
gatgcaaata acctagggtt tttgtattta attttcattc ctttggattt tttgtttcct 9360
cacgaagttt gaataaaatt accaaatgtg gagtacacca agaagacagg tataaatgta 9420
ggaatgaata aacttatgta tgtatacatg tatggcagag agaaatagag aatatgtatg 9480
tttgtgtaag ttatgtgggt ttgatgtata gaaagataca gattaaaaca gacatatagg 9540
gagacaatgt tatgtaaaat ttccgatgtg attattgaaa caagagaagt aattgtcacc 9600
tagataaata gatgaatgag cgaatgataa atggatgaaa caaatgccaa atctgaatca 9660
gagagaaatc ctcacattct ttgtcacttt cagtttcaag agataagaag atgttctccc 9720
caaaccacac catagtgaca gaattcattc tcttgggact gacagacgac ccagtgctag 9780
agaagatcct gtttggggta ttccttgcga tctacctaat cacactggca ggcaacctgt 9840
gcatgatcct gctgatcagg accaattccc acctgcaaac acccatgtat ttcttccttg 9900
gccacctctc ctttgtagac atttgctatt cttccaatgt tactccaaat atgctgcaca 9960
atttcctctc agaacagaag accatctcct acgctggatg cttcacacag tgtcttctct 10020
tcatcgccct agtgatcact gagttttact tccttgcttc aatggcattg gatcgctatg 10080
tagccatttg cagcccttta cattacagtt ccaggatgtc caagaacatt tgcatctctc 10140
tggtcactgt gccttacatg tatggcttcc ttaatgggct ctctcagaca ctgctgacct 10200
ttcacttatc cttctgtggc tcccttgaaa tcaatcattt ctactgcgct gatcctcctc 10260
ttatcatgct ggcctgctct gacacccgtg tcaaaaagat ggcaatgttt gtagttgcag 10320
gctttactct ctcaagctct ctcttcatca ttcttctgtc ctatcttttc atttttgcag 10380
cgatcttcag gatccgttct gctgaaggca ggcacaaagc cttttctacg tgtgcttccc 10440
acctgacaat agtcactttg ttttatggaa ccctcttctg catgtacgta aggcctccat 10500
cagagaagtc tgtagaggag tccaaaataa ctgcagtctt ttatactttt ttgaccccaa 10560
tgctgaaccc attgatctat agcctacgga acacagatgt aatccttgcc atgcaacaaa 10620
tgattagggg aaaatccttt cataaaattg cagtttaggc ttgtgtttat ttgcagtcac 10680
gaattgcttg tggagtaaca aactggcttt tgaaatggaa aaacctagtg tagtcgtgat 10740
ttatttaaca tcatggactg tcagtaacca ctttactttc ttatccaaat gaaaaccttg 10800
aagattgatt tcttagaaat aaaagcctta atgttgagaa atttaaaatg ttttatttgt 10860
cagaaattct atgaaaataa atttttttagt atctaataat tctatatgaa aatactatgt 10920

```
ttacttactt gggggggtggg taaatttttа actaacttag cctgagagaa attctgaaaa  10980
ttcatctcct ttgaatatga tggtccacag agccataatg aggttcaaga tagggtggat  11040
aatttttacac ttttttaaaaa atgaaattat ttctacttca atatggtgtc ttgaacaata  11100
gctcagctac ctttgtgctt attgtaatca atgtgcaaaa aatataaata tatgggtcac  11160
tgatacgtta aatttgctgt aagtgttccc caaacccatc ttgtaaatat cttgagaagc  11220
ttctttttтg ttcatgtgat aatactggtt atcgtccttg attctgaaaa tcaactctca  11280
acacctttтt atgagtttcc aagtgtatgg atccaattct atcatcaaac tactttaact  11340
gcattccagt tatgttttcc agtgccgtac ctctgactta ggactaattc agtaggagtc  11400
aatagggtct attaactgcc ctgtgcatac ccataaatgt cttgaaacag tatacaaata  11460
aataggggtg tatatatgtt tctggaggaa atactcagtg tttattttgc ttcatgaagc  11520
tttctgtggc ccattatttt taaaataagg gtgttctagt agttcttgaa gctgctatag  11580
agaaaatcaa tctcagagag gctacctaag gagacacatt aatccatgca ttattagtca  11640
accagatgac tgcatattga tacatatgta tttagcccta ataaattgcc tcatatgttt  11700
ggcattttaa tattttgcac tctgatatta actaacacca tacagaatat gctacagaaa  11760
atgtctgtat gatatcaact actatatcta ctataaagtg ttcaaatggt tttccagaat  11820
attaagcaga tgacttgcag tttttgtaaa atgggggcat attcttgagt ctcatttgat  11880
taagtctggt tctagccaaa tcctctttca gaaacagtat ctgaaaaaga aggaaggcaa  11940
atttcggtga caagaattta acaattacat tttagcactg tttgtttttg tatttgcttg  12000
gttttggcaa tgatggtggg acaacagagg aaggtcatct gcagagacaa caatgaagca  12060
gaaaagctgg gaattgagtg acacagagtc ctcacttcct ttttgagggc cagcatcatt  12120
tcatcggtgg ggtatttctt tattttatga gaaatttcca tctattaggc aaagcattgt  12180
attccatttc tggcaagcag aatggtctca tatccagatt tgaaacccta tttgaaaatt  12240
ggattttaga atgagtgcaa aaaatactct atatttggcc taaataactg tacgattcca  12300
tttatttttct agttatgaga attgctctgg cttgcagcaa ccatttttaa actttgacat  12360
ccttttagtc aaaatgactg acctgtttct aattgatttt taataaaata ttttatttat  12420
ttctatttgt gatttaattt tttccttctg catgtctttg ggtaaaagtc acctcagaaa  12480
ttctttggag ataggataga acataaatcg acatgaataa aagtaattat aagagttgca  12540
gtcacagtgt agtgctgatt ttaggaacac tttatttttaa tcttcataaa tgaaataatg  12600
agagaattat tatctccatt tacaacattg gagcctgagg cacggagaga ttaaatttct  12660
tgcccaaagt catttggcca ctcagtagga caaaagcagg tttcttggtt tcaaattctc  12720
tctctttcta attcatactg tagttgttat taaagaaatt ctcaaaacac tgcaaatgct  12780
tggtggaatg ttggagatta tgaattgatg agattattgc tatagatgaa atagtggtta  12840
ttatgtgata gttatatgta tcctaagaat agctaatgag aatttccctt gaaaattccg  12900
gttaacataa tttataggca ttataattaa aattgtctta acgtccatgt gaagggagaa  12960
aattttgata tacagaagac atatatgcta tttaggctgc attttattaa tatattttcc  13020
cagtcgcaaa ggcatacata ttcatattaa ttaattacca ggtacagaca aatatgaggg  13080
agtaaaattt atgagggatc tgtcctgtga caaaggatac attttcttta tattttgtta  13140
catttcttct caataagttt atatgcgtac atgttactaa aatgcatcaa gcaaatgtga  13200
aatcacttat catatacaaa tttgaacact gattttcact tgatataatg tatgcacctt  13260
tatatagatt aaatatttct gatgaacatc atgttgaatg tcatataata gttagatcat  13320
atgaccttaa tacaataata atacatatca attgtagaaa actgactgaa ccatattgac  13380
atacataaga catagaagca acatttactt ataatcacct acctgaagag agctgaactc  13440
agcattcact tgtgttttat tccatttgat tatcttgcaa accatgattc catcatatca  13500
atgtatctta gtgtttttat gactgatgta tctgtgcatc tctaattgtt ccttttgtgt  13560
atgttcccag aagtggaact gctgaattaa atgttttcaa tgcttttgaa actcctgcta  13620
aatttttcta tgggtaaatt tttctccaga aagcttctag caacttaaac ctgtactaga  13680
aatatttgag aataatcatt tcactataac tctatcagta gtaggagctc ttgtttatgg  13740
ttatttgctt ctttgatgtc tcaagggaag tcttagaaaa taattctgtt ctaataagca  13800
cagaataaaa attgaagcat atattgaaca aatgaaggct acacataagg ttgataattt  13860
ctcttcatgc tattgctgat aatttcttgc atgaataatg tttgtatttc tatttggaaa  13920
aatgtgacct tttgcttatt gctgttatgt aagttttcct catcctaaaa ttgaaaaaac  13980
aagcacttat attcattgca atatttgata gtgatatcag gaattatttt agatataata  14040
cattttgta ctgtgttgta atattaaaaa ttatcttttc cctcttattt ttaaccaatt  14100
tcctcccact atttattaag caatgctttg tttttcaact gatttgtgat gtcaccttta  14160
ttataccata agtgaatcta tttctgccta tttcctcttt atgtttggcc acaatataca  14220
tatattttaa actaaattcg tatgattagt gtacacttaa aatatgtttt aatttctgat  14280
ggaaaaatgt tatcttcact tttctcttag gagactgcaa atcatctgat gattgctttt  14340
agatatccag tatttttaac ccataaatta tgcatactat tgttcaattt agaaactata  14400
tgcatacttt atatatgcaa atatatattt caatgtggct atacatgaaa tacttaaaca  14460
ctgtaaatta tttgaaaaac aawgaaaact attttcctct ctatgggggc cataggaaaa  14520
ctgcttattg ttaagtttgt tttccagaaa ggcaggaaaa cgttggtggc ctgtgagtaa  14580
ggagaaaaca cttaaccttg gtatattaat tctatcgatt cattaacaac tttatacaaa  14640
catttatttt tataccatac cggcgttcat aataaaacac ttttttttcaa caaatatctc  14700
```

taaatacaga ttcttttctc taactatgct gacttttgc caacaactcc acccaaaaga 14760
gttgacatgc ccacatatat tcttctgcta ggaatcattt caaacactga aaaaacccta 14820
ttaattagaa ttaaaataaa tattaaacat ttaaataatc agagaaaaat tactctgtgc 14880
ttcctaagaa tgggaatact tcacagtaat actgttgctt taatttgaca ataaaactat 14940
ttaaaaataa gacccaccct acagtctggc aatgtatttg tcacaaataa gacaaagtgg 15000
cacaattact gaaatatgtt atgttaatat aaatcatttg aagcactaaa acatgaccat 15060
ataaataaat gttgaatact atacgtagaa aattgattaa aatacaaaat gctaaaatta 15120
acgcagaaaa tagtggcagc taaaaacatg aagtacaata ctgtttatgt gtctatattt 15180
aagacagctg aaaaagtcta gcacgttcat acattggttg attaatggta taaaattttt 15240
ctagaaataa atacaatatt attaatattt taaggagtca aatgtttcta taattttatt 15300
cagctattcc aaatttaaga atctactgaa ggcataatta gaaacttgtg tgaagagttt 15360
catggtgttc attttataat caagaaaaag caaacaaatg atgttcgaaa tacaagtagc 15420
tcccacttag caccgtacca tgtaacagaa acctttctat atggcaaacg tgtcttgttt 15480
tgcatggttc ttacttagat attacaaaat ctcaaagtga ggactgcctg ctttggataa 15540
ctggtaaagc acaggcatct accagatggg gtagaaggtg tttctaaata actagaatgt 15600
catcagcctt gatcatgaag gaaagtgtaa agtacagttt aatttttttt tcttagaatg 15660
ttatacagct attaaagtga ggatgttcct agttaagaaa aaggtaaatc attttgtgta 15720
atatattatg tgcaaagaca tatccaatgc tacatatgga tatcttgtaa aatatcaggg 15780
ataaagagaa acaaaactta ttaatagcta tgtccaaatt ttccatgata agcaattgtt 15840
gctattacat tgagaaaata agctaacaaa aactttgtat gcccagttta attcagatct 15900
caggatatat aattacagga tgcttattta taatagtaga atttaaaaat aaaagtagga 15960
aaaagcaagg agttaaaaac ttgttgtaga gcgtgatggc agaagaggag gtgctcagtt 16020
ctggtttcca gaacagaatg accagttagc aacttctcac agataagaat gcctaggtaa 16080
aaattccata acccaagggt gaggtggggc acactcttgg agcacagaaa ctgggaaaag 16140
tcatattaga agtgtaaggg gagcagtttc actttgactg tgttgcttct ttcccgacca 16200
aacagtgtca cactgaaagg gatttcttgg gcctgcgttc tctagtggag aaaagagagc 16260
ccacggcaga catccaactt ccctgtgttc cagggcactt cccaaggggc ctagttctgt 16320
ctaatcttgt ggggaataat ggaggaattg gcagggcttg acccccttatg gtcagttcat 16380
aactacctcc ttctacaacc cattctgtat tccctttacc ttcagcaagc accttagcag 16440
gacacggttt tttacctggt ggagtgacac aaatcttcat tcctgatggg tctgggccat 16500
ttgtagtttt gcctggattg ggttgttgta gtttctcact gaccttaatc acaggacatg 16560
gtaatactgt gacgtcctat gggatctcct gtattccaca aatactcttc cttaacctcc 16620
attgtggagt agtagtttga tttcatcttg atagtctggg tcagtcaccc cagccaacac 16680
tgtacctccc ttcttagcct gttgacttag aggcaggaag agctcaaagt ggccaggtgg 16740
caaacttatc ttccagtttg atggaatcat ctttctgtct tctggtggca aagttcctcc 16800
ctccggaacg aagacctcta ggcccacaga acataatgtc acagaaacag gaaggaaaat 16860
tttgctagtg ggtcagtaga ggtgatggta agtggcacca cttccacttc cacccccttga 16920
ttcctgcatt cttggatcct ggctatggaa gaaacagtac cacatattgg atgcctactc 16980
agagcataca aagacttctg gagaaatttg ccccaggcct gcatttttttt ttttttttt 17040
tgtaaaatga agtcctgctg tgtcttccag aggcctatag tgcagtgtca cgatcttggg 17100
tcactgcaac ctctgcctcc caggttcatg cgattctctt gtctcagcct cccgaatagc 17160
tgggattaca ggcatgtgcc accacggctg gctaactttt gtatttttag tatggatgag 17220
ggttcaccat gttggctagg ctggtctcga actcctgacc tcaagtgatc cacccgcttc 17280
agcctcccaa agtgctggga ttacaggcat gagccaccgt gcccagcccc tgtaataaag 17340
tatcacctag ttgttgtaac tgttacttca aaaggctttt ccacctttct atcaatccat 17400
ctgctttagg ataatgggaa acatggtaag acaagtatat cccatgagca tgagcccact 17460
gtcttattct ttagctgtaa aatgagtgcc ttggtcagag gcaatggtct gtggaatacc 17520
atgacagtgg ataaggcatt ctgggagtcc acgaatggta atcttggcag aagcattgca 17580
tgagagatag gtaaacctat atccagagga agtgtctaat gtgactaatc cactctagaa 17640
ttctaatcac ccaaagcttt tggattcctc cctctacatt aaaccaggga agatctggca 17700
tttcaaggtc gctcacagta ggccatcctt tgatccatat ttcagctaat aaactattag 17760
aacatttttt aactccctga tttgtaacat tgagtgcaga atctctgctt agtgggccca 17820
tatcaataca ttcagcctga tccaacttta tattcctccc accattatcc cacaccctta 17880
atatccattc ctacacctgt tctccagaat tctgcttata taaactagaa aactcaagca 17940
gctcttttag ggtgtagcac acctccttgt agacttttgg cttacaccaa ctagcaccag 18000
atcttggaag gagaaaaggg attaaagtct gaatgtattt cctaaagaag aagaccaacc 18060
aaaataggtt atttaaatat ttggctattc taagtccaaa gctactcttt aaaaggaagc 18120
catatttaac cactccatgt aatatgtaat tatgtcttga catacagaac taggtaatgc 18180
aataattaaa ggtatttggt caccataatt tttttctaaa aacggagaaa tgtgcaaatt 18240
atttacctac cagtttacac aaacaggcaa agcaaaaaca aaagcacttg tcaaggtgga 18300
caaaaatgct gctacagtta aaaatctgca tatactcttt actttgtcat tcttaaaaat 18360
gtgtttctgg gtagttggtt tactgaaata ggtgaaagct gtactgtaat gtcattttta 18420
aattttgtaa cacttttttgt aacattttgt aacatttggt gaagaaatta tgtttcttca 18480

cttaaggaga taaattgtgt tcattataga actttgaaag tgaacgacca aaccaaatgt 18540
gcaaaaattg actgaatatg aattacttta cacttatgtg tgaaaataat cttcaacatt 18600
atgtattcca tacacatgtt tttggtgttt taaattatct ccgggctttc ctaaattcaa 18660
gttggagaag cagaattagt tagatttata gcacactcac tatagaaaac atattttgaa 18720
aacttgtttt accacatctc tgttcagctg gatggggctc ctgtgaaaga acaaagcaat 18780
tacatagaat tttacacaga attgtaaaac ctgcaaattt tttttaacat tttctttacc 18840
ttatctgata gcacagcaca tacgttgcaa ctcttactct gaaaactact atcacttctg 18900
aatgtttatt aaataacagc cataggtgtg cgaaggattt acaaatggat agccacattt 18960
actgccaatg aaaaacaaag tttatttaac atgtgcagta tctttgtgat atatatatta 19020
ctccttattt ggtacagatg aaaacgctga agctcagaag aatgaaataa tgttttccag 19080
gaacacagca agagctaggt ttcaattaaa atattattct cgtagctact tctgggttga 19140
ctctagaaac accccaagcc acatgggttt gttaagatgt ttaactgggt aattattttt 19200
atgggtaatt atgtttcaga cagttttaca attttacaga tataaaaata tatatgaaaa 19260
agatatatat atgttgtaat gcatatatat atattcatat ttcatatgaa gaaggcacaa 19320
gaggaacact cagtgcctag gtaaaaagaa tggagttgga tagtaggtta atagtggttg 19380
atgagataag aagtagtcat ggtcatgact tactttcctg gtgtcatgca atttctattg 19440
acagcatttc ttcaaagttg taatttttat tgagaataaa ttaataaaac tctgtaagat 19500
tttaagataa aagatctccc aatattttca aagcatcatt acaaccttct ttaaagtttc 19560
tagcaattgt atcaatttct ttattcttat tttcattgta gaaaaaaatt gatttttgag 19620
gtatyttaaa agtcatagag aaaatatttg gccatggaaa gggtccaaaa tgggaggagg 19680
aaattaaaaa ggactgagat ttctgaaaca catacattgt tccaagggag taccaagcat 19740
gctacaaagg aggcagagga ttgaatgact gaagaaaaga ttttgaagga cactttgagg 19800
tgttcttcct acaccttcca tcatcctatc tcaggagggt gaatggattg gagagattat 19860
aaaaagagtt attgtgatgg ccttgatgtt agtccaaaac atccttttct agtctatgtg 19920
attgataagg atgctgaagc tgttcatggc tgtggaaagg gaaggggaga ggagagttat 19980
cccattgctg tactgtatct gagtgacttc acctcagagg aaaggaacag agaagcatct 20040
ctaaaccaca ttctcaaatc caaattcacc agcaataatg ctagtaattt tcatatctga 20100
ctcccatgtg tagactctaa ttgagtcagg atccaaaagg agggcaattc aatggctcct 20160
ttaagttcta acatgatgaa tgtagaagta ttccaaattt ttagctaaaa taatatttat 20220
cataatagtt attattactg ctactaactc actgcaagac ttatggaata actactatat 20280
gatgctctgg ccaggatctg agatgtccac aattctactc atatgtttta gctaaagaga 20340
atttaatgaa aagacagaga ataaaaaggt aaagttaaag gaatcagcca aaggggttaa 20400
aaaagtcaag gctaactgta gaaaaccatt tccatcttta ggtctgaagg accaaggaca 20460
gtggcttctg tcacagagct cagtgagagc tgaagccagg aagaaatccc ttgccaacat 20520
cttcccaaga cacacatccc aagacattct ccagggaggg atgcagccaa tttagcacac 20580
ttrtaatgaa gcaggaaggg aatgagaaag aaattgaggt aattccctat ctttcccact 20640
tctgacatcc tagagagtct tctcaaaagc caaggtcagc tggaaaccag acaccatagg 20700
tgccctgttg accaggtcca taggagccca actccaggtg aacaatgcaa ggcacagaag 20760
agaagacaat agatgtgaat gggagtaggg tggaaaagga gaatcaccag cacacctata 20820
ttatctataa atccaaaatg ttccacactt acacattact atccctatta aaaagataaa 20880
tgtacaaaac tcaaaattga tatatcacta gttcaagttc aacaaaagtt agtttatctt 20940
ccaatggtgt tccagactct gaaaccttct cccgatacca tccactatac ttgctgtaag 21000
caggaactta atatttccca aagtctaaca tattaaatta tttgggagta aagtaaaata 21060
aataatgaat cccattccta catcactctt gggggatgcg attccctgaa gacagtgtgg 21120
agttggcaaa actcatgtgc tccgtgagct taggattttc agaagaatcg aaatattgtg 21180
gttgttcata atttgaatga ctttttgcta gaaataattg acctgcaggg atttggtctc 21240
cacacagtaa ttgcattata aactctacag gttctcaata cccattctgt aatcatgaaa 21300
tcatgcttct ttctcaggta ggattagata tatttttaa aaagtttaaa atccggcagt 21360
aaatatttta attgatagaa ataaatggag ctgaaacata ttctgatgaa ttagaaaatg 21420
tttatattaa agtgttgcta ttctttcag agtactggtg tttacagagt gtttgcataa 21480
tgaaaagaaa ataggccaaa agaatttta tttcatcaat aattattgac tgcctccaat 21540
tcaggaggca caaatctagg tgctgctgag tctataacta gatcagataa aaatccttga 21600
ccctaaagga gctcatgctt tgatgttgag tttcgttcta gactagacct ggatcattca 21660
ccaaataata ctagtagcct caatttgtta agcatgtagt acatgacaga tatattttaa 21720
aaacttttac tgaattaact ttctcatcaa gagagaatgg ctattaatca aattttatgg 21780
ataagaaaac tgaggcacac agaaattaat aacttgcccc aggccacaca ggtaggaact 21840
agaggcactg ggtttttcac ccaggaagtg tgactgcaaa ctcccagtaa ctgtcccact 21900
ctgctggctt tcaatagtaa actatgctta ctgtttatac ttgatttaga acaatcaaga 21960
tctggtatta aatcatcctt ctgattcatt ttgtgttaac aatgccactc tttactatta 22020
taagagctgt atcatgtaat attttatagc cctttcataa agaatgggct acttttaggt 22080
ttatagccta tattacttaa aaataataaa aatatgtttt gtgaagctgc cttctctaga 22140
tttctttttc cactccaaat tgaaccatga ccactttcca aggagacacc actcttgtta 22200
aagacaaaga ccaaccttct agactctttg taggacagaa tagaagaagc aagcattcac 22260

atgggaatgt ctcagtgtag ggtgcatgtc aggagagcac aggctcagat gcctccaaag 22320
ccatttctgt ccaattgcag aagcttctcc ctgagatttc cttgctcaaa gcagtaggat 22380
attgtgttct tgtttttca gaagaaagac tatttccaca aaattctcag ttaaattatt 22440
tggactattt attgcttgca ttacttcaac ttccaataga aaaaaaaaaa gactcataaa 22500
aaaacagatg aattgagaag ttattttgaa gggtgattaa tatttcaata aaaaagccca 22560
ttaaaaatgt aattaggcgg ggcacggtgg ctcccgcctg ttatcccagt actttgggag 22620
gctgaagcgg gtggattgcc tgaggacagg agttcaagac cagcttggcc aacatagtga 22680
aaccccgttt ctactaaaaa tacaaaaaaa ttagctgaac gtggtggtgg gcacctgtaa 22740
tcccagctac tcgggaggtt gaggcagcag tattgcttga acctgggaga cagaggttgc 22800
agtgagctga gattgcacca ttgcactcca gcctgggcga caagagtgaa actccgccac 22860
acacacacac acaaatgaaa tccagcacca aaacagtatt tgcctatttg gcaattaaat 22920
gtaccatcaa acagagggat agagaatgtg gtatattata tatatatcat acattataat 22980
gtgatataat atataataca ccacagtttc tttatccact tgttgtttga tggatatata 23040
catatatatt atatgtaata tatatacaaa atgtatatga tggaatacta ctcagccata 23100
aaaagggatg aagtaatggc attcacagca acttggatgg gattggagac taatattcta 23160
agtgaagtaa ttcagaaata aaataccaaa tatagtgtgt tctcactcat aagtggaaac 23220
taagctgtca ggatgcaaag gcataagaat gacacaaagg acttcgggga cttgggggaa 23280
acggtggagg aggtgaggga taaaaggcta cacattgggt tcattgtata ctgtttggat 23340
gatgggtgca ccaaaatctc acaaatcacc actaaagaac ttactcatgt aaccaaatac 23400
caccttttcc cccaaaacct atggaaataa aaaataaaaa taaaaaccaa aagaaattca 23460
attggaccct gtgagcttta acaaggtaat tatgattcta ttgatttagg tgacttatct 23520
tctaacttat taccctaggc agaagcccaa tgtccttttg tgtagaatga gataatacgg 23580
tggacaagtt ttgaacattg aatttacaga gtgctttatt tatgaagtga cctgtttcca 23640
gtgttggcat ggtagttatt aaaaaaaagt tagtataaca ttgtcaagtc tggatatgtg 23700
tcatagaaga aacacgtcga ttgttccatc atgatcttca tggtcacttt tatcttgccc 23760
aaatgttgag agattcaggt caggatattt taagtatggt ctcagtcatg ttgttaatat 23820
gattatgatt tcactttgat ctgttacatg ttttaacaaa ttcatattga acttgtgttc 23880
caatgttttt tcttagtttt tcaagtgtta gaatgaggta gagttaaaac aacagcattt 23940
tagtttgagg ataagtttct tcatcatttt atccttattg ccccccctccc caggagccct 24000
ttgatacact attgttttct ttatatccct atcaacattt aataccatca atacaggtgc 24060
acttgtgtat ctatatacca tgcaaatgta tatggtttat gtattatgtg catatgtatt 24120
ataaatacaa atatgtaata tctactttca ttccccaaga accaattttt gccccccttgg 24180
acaaaaatat acaaaatgaa gaaatgtagc ttggtcttta aagttaaaaa agatttgttg 24240
ataagagggc cagctgaggt agaataaagt gaagaaaatg ggcacacatt agagggtgag 24300
ggaagatctg gaaaatgatc tccagttcac aggggcaggc aaagcgattc ttgttctaca 24360
gggatgtata ccaggcatca gtcccacttt cctaagcacg ttcagttgtg ataaacctgg 24420
agggtttcaa aggctgatac tttagatccc acattcaaag gtgtgttgtt aaacaaagaa 24480
ttacagtttc aaagaaaagc aatgtttaca accatgggtt caagaaaagt ctaagtgaac 24540
acatataaca aagacttgca aaaagataaa agataaggct ctttaactat caaaagactt 24600
gcagaaaaga accacagaaa accattttaa atattattgc ctttgtatat taaaaaactc 24660
tatattagtt tagatgttta aagcatcaat cacatgctca ctaggctatt tcttaatgtc 24720
acatgtattt acattttgag agaagaggaa gaaatagcag atgacaccac tggggtaatg 24780
cataaatgac aaacctaaat gcattttaat ttccttttat ttagatgtca tttgaagcca 24840
agcaaacaca atgttaaaga aaaaccatac agccgtgact gagtttgttc tcctgggact 24900
gacagatcgg gctgagctgc agtcccttct ttttgtggta tttctagtca tctaccttat 24960
cacagtaatc ggcaatgtga gcatgatctt gttaatcaga agtgactcga cactacacac 25020
tccaatgtac ttcttcctca gtcacctctc ctttgtagat ctctgttata ccaccaatgt 25080
tactcctcag atgctggtta acttttttatc caagagaaaa accatttcct tcatcggctg 25140
ctttatccaa tttcactttt tcattgcact ggtgattaca gattattata tgctcacagt 25200
gatggcttat gaccgctaca tggccatctg caagcccttg ttatatggaa gcaaaatgac 25260
caggtgtgtc tgcctctgtc tggctgctgc tccctatatt tatggctttg caaatggtct 25320
aagcacagac caccctgatg cttcgtctgt ccttctgtgg acccaatgac atcaaccact 25380
tttactgtgc ggacccaccc ctcttagtcc tcgcctgctc agatacttat gtcaaagaga 25440
ccgccatgtt ggtggtggct ggttccaacc tcatttgctc tctcaccgtc atcctcattt 25500
cctacacttt catcttcact gccattctgc gtatccacac tgctgagggg aggcgcaagg 25560
ccttctccac ctgcgggtct catgtgaccg ctgtcactgt cttctatggg acactgttct 25620
gcatgtacct gaggcccccc tctgagacat ctatacaaca ggggaaaatt gtagctgttt 25680
tttatatctt tgtgagtccg atgttaaacc cattgatcta cagcctgagg aataaagacg 25740
ttaaaagaag tataaggaaa gttattcaaa agaaactgtt tgctaagtaa ggtagatatt 25800
ttggtcatag gcgttggaat ctgttcttat tatctgacca attaatgaac atttaaaatt 25860
aacaaatcaa tctgtcattg agtgtttttt gtcctttgta atttgcatat gggacttaaa 25920
agtgtatgtc aaattattag ctagagctta cactgccacc tcagtaaatt gaaaatgaaa 25980
gcatagaaat tcaaatataa gactagaaga cattgttcta gctctgtaaa aagtaatgaa 26040

gtacagatga ttccattatt agtactcata gtactaatat atcatagtac tatatcatag 26100
tactaatata tcatagtact atatcatagt actaatatat catagtacta tatcatagta 26160
ctaatatatc atagtactat atcatagtac taatatatca tagtactata tcatagtact 26220
aatatatcat agtactatat catagtacta atatatcata gtactatatc atagtactaa 26280
tatatcatag tactatatca tagtactgta tcatagtact aatatatcat agtactgtat 26340
catagtacta atatatcata gtactgtatc atagtactaa tatatcatag tactgtatca 26400
tagtactata tatcatagta ctatatatca tagtactata tatcatagta ctatatatca 26460
tagtactata tatcatacta aaatatcata aaagacagaa ttatcaattg ctgagaaaat 26520
gacagctcat tacatttttt taaatgtttt tgtttgtccc tttgaatgaa tggaggatat 26580
ctacctaact tctatgtatt cagaatctca catgaggcag cacatttagc acacttgttt 26640
agttgtaatc ttctatgctt atcttcgtga gtctgtttga aaagttggat aaatggcaat 26700
gagtgactac atttgctaat tacattttat aatacttaat aagcctccaa gctattttcc 26760
agaaacttgc caaccatagt agatgaggtt tctgtttttc ttttatacat attatataga 26820
aatgaagatc ctttaggtgt actttggatt attactttaa gttatatcct tagaatgata 26880
tcatagatcc aatgaaaccc tatttcataa ggcttttcat ggaaatgcca aatttccttt 26940
agaaagagtt tttaattgtt tgtatgttgt accagtaagg attcccatca acatgtgaac 27000
tcaattcttc actcttttat taccattgcc tgtttccttt gaataaataa attgaattga 27060
atgttttatc cacgaatgca tgtggatatt tgtgtgtttt atgtgtgtat gtaggtgtgt 27120
gtgtgtattt tgattacaat ataacatttt ctttttttta attgcttatt caaatttttt 27180
tggtgagaat caaggtcgta tgccttttt ttttttttt ttttttttga cagagtattg 27240
ctctgttgct caggctggag tacagtggcg cgatctcggc tcactgaaag ctccgcctcc 27300
cgggttcacg ccattctcct gcctcagcct cccgagtagc tgggactaca ggcgcccgcc 27360
accacgcccg gctaattttt tgtatttta gtagagacgg ggtttcaccg tgttagccag 27420
gatggtcttg atctcctgac ctcgtgatcc gcccgcctcg gcctcccaaa gtactgggat 27480
tacaggcatg agccaccgcg cccagctgct tcacctgttt ttaaattgtg gtgtttatat 27540
gttttcttat ttatttcaga tatttctcaa tgatataaac atgtttttgt tatatataca 27600
ggtgtagcaa ttgtatatat atacacatga attccatttt atctcttcta ttgtgtcttt 27660
tccaaagtta atattttatc attttaaatt ttctcaatat ttttttattt cactttattg 27720
tgtctttttt tacttttagg aagttaagtt tatgccacct atgatcatgt ggatttttac 27780
atatattatt aaattatttt catggtttcg cttgttaatt ttttcttatg aactattttt 27840
aatatacatc taaaatagaa aataatataa agactccttt gtaagcacaa ctaatcacta 27900
ttaaatcttc acagaatgcc ccatttgttt cagttatttc agaaatttaa aacttacaaa 27960
gatacatctc tcctgactga ttctgtctct tgtcccattt cttgctgtaa actctattac 28020
aaatttaatg tttcttatcc ttctgtagag ttacatcatt ggataggtgt ttatatattc 28080
atagaaatat gtagtactga tttgctacat ttaaaacttg aaataaatgc atccactcta 28140
tgtatctttc tgcaatttat ttatcttgca ctaaatattg ttggatgtat ctgtggtgaa 28200
tcatgcagca gcaattttta gtggtatagt ctatgccaat atataaatat gcctcaatat 28260
gtgtatttat tctacgattt ttttttttt ttgagacaga gtctcactct tgttgcccag 28320
gctggggtgc aatggcacga tctcggctca ccgcaaactc cgccccccag ggtcaagcga 28380
ttctcctgtc tagcctccta gtagctggga ttacaggtgc ccaccaccat gccaggctaa 28440
ttttgtattt ttaatagaga cgggatttca ccatgttagc caggatggtc ttgatcacct 28500
gaccttgtga tcctcccgcc tcagcctccc aaattgtatt ttatttctac tgttggtgaa 28560
gagttacatt ttccagtttc ttgataatat gaaaatccat taataaatgt tttgtgtgta 28620
attatgaatg tgtaccacgt aaggagttcc tctcaggtag gttggtaatt tcttggtcat 28680
aagatataag tatctttaaa tttactatat attgccaatt gttttactga tttatatttc 28740
tatcaacagt ttgtgagtgt tacctttaca tcactttatt cttgccaaaa ctaggtgttg 28800
tcagacttta aatgttttct aatctgatgg gtaaacaatg aattctcatt tggtgtttaa 28860
ggttgaattg ttattacaag tgaaactaag tatttttaag tatattgatg ctaatctcat 28920
atttctcttt cacttatatg tccatatact ttgttaactt ttcaggtagg gtgatttgtc 28980
tttatttcat acatttataa tagttcctta tgtaccaaat gctacataca aatcttttgg 29040
tgttatatgc catttaacct taaaaaccta gggtaataat gccaggcagt gtggctcaca 29100
catgtaatcc cagtactttg ggaggctgag gcaggaggat tgcttgagcc caggagtttt 29160
aggccaaact acacaagata gagactctgt ctttacaaaa aaataaaaaa aaattagcca 29220
gttgtggtgg gatgcccttg tggtcccagc tacatgggaa gctgaggcag gaggatcact 29280
tgagcctggg aggtagaggc ttcagtgaac catgttgtca ccacgacact ccagtctgga 29340
tgacagagga tggatgacag agcctggatg actccagcct ggatgaccct gtctcaataa 29400
aacaaaacaa accccaaaaa acctagggtg agatgaacag atgaacgtta tttcttgaat 29460
gcctcagcaa atagccatag gtatctattt ttctggtagc tgcaatttga ttgggacaag 29520
tttggaattt cagctcattt gacaagtatt ttcataatag atattttccc agtaggattg 29580
gatttttttt ttccacatag tgaatgctta tatgtgagca taaaatgttt cccacagatt 29640
gggattggat aaattatatg aagtgtgatc ctagtatatt gctttgccat tgaaatactg 29700
gttattaccc aaatagaaaa ttaacattta aaagagccaa caggctgggc aaagtggctt 29760
aggcctgtaa tcccagcctt ttgggaggcc aaggcccaca gatcacttga gttcaggagt 29820

tcgagactaa cctgggcaac atgaggaaac cctgtctcta aaaaaaccac aaaaattagt 29880
taggtctggt ggcaagtgtc tgtagtccaa gctactcagg aggctgatgt gggaggatgg 29940
cttcagcctg ggaggcaaag gttgcagtga gctgagatca tgccagtgca ctccagcctg 30000
ggtaacagag taagatcctg tcttaaaaaa ataataaaat aggccaggtg cagtagttca 30060
tgcctgtaat cccagcactt tgggaggccg aggcaggtga atcacaaggt caggagttcg 30120
agaccagcct ggccaagatg gcgaaacccc gtctctacta aaaatacaaa aattagccgg 30180
gtgcagtggt gggtgcctgt aattccagct acttgggagg ctgaggcagg agaatcgctt 30240
gaacccgggg ggcagaggtt gccatgagcc aagatcgtgc cactgtactc tagccttggt 30300
gacagagcca gactccaact caaaaataaa taaataaata aataaaataa taaaatataa 30360
tataataaga atgccaacag aataattctg atttcagcta cagtttatgc ttttttatga 30420
gaaattcaac tccaatcttt taatttttatt ctcctttatt ctgttgttca gacattgcat 30480
attaaaaagg catttttttt cttaaccaat agttttatcc ttattgctgt ccttattctg 30540
gagagaactt catatttcaa ggcacagccc aaatatcaac tcccttgtcc tttcatgagc 30600
aaacagttgc acatacccat tcctgtgatt gtgtagatat tgcaactcta tttgtctgtt 30660
agtcacacag aaaattcata ataatacaat gccatgtttc ctaatatatt ctgagagttc 30720
tgtgagaaat gaaacctatt gtatcggtta agtactgaag atggccgaat aggaacagct 30780
ccagtctaca gctcccagcg tgagcgacgc agaagatggg tgatttctgc atttccatct 30840
gaggtaccgg gttcatctca ctagggagtg gcagacagtg ggccgaggtc agtgggtgcg 30900
cgcaccctgc gcgagccgaa gcagggtgag gcattgcctc actctggaag cacaaggggt 30960
cagggagttc cctttcctag tcaaagaaag gggtgacaga tggcacctgg aaaatcgggt 31020
cactcccacc ccaatactgt gcttttccaa cgggcttaaa aaatggcaca ccaggagatt 31080
atatcccgca cctggctcag agagtcctat gcccacggag tctcactgat tgctagcaca 31140
gcagtctgag atcaaactgc aaggcggcag cgaggctggg ggaggggtgc ccaccattgc 31200
ccaggcatgc ttaggtaaac aaagcagcca ggaagctcga actgggtgga gcccaccaca 31260
gctcaaggag gcctgcctgc ctctgtaggc tccacctctg ggggcaggac acagacaaac 31320
aaaagacagc agtaacctct gcggacttaa atgtccctga cagctttgaa gagagcagtg 31380
gttctcccag cacgcagctg gagatctgag aaagggcaga cttcctcctc aagtgggtcc 31440
ctgacccctg acccccgagc agcataactg ggaggcaccc cccagcatgg gcagactgac 31500
acctcacacg gccgggtact ccaacagacc tgcagctaag ggtcctgtct gttagaagga 31560
aaactaacaa acagaaagga catccacacc gaaaacccat ctgtacatca ccatcatcaa 31620
agaccaaaag cagataaaac cacaaagata gggaaaaaac agagcagaaa aactggaaac 31680
tctaaaaagc agagcacctc tcctcctcca aaggaaagca gttcctcacc agcaacggaa 31740
caaagctgga tggagaacga ctttgacgag atgagagaag aaggcttcag atgatcaaat 31800
tactccaagc tacgggaggt cattcaaacc aaaggcaaag aagttgaaaa ctttgaaaaa 31860
aatttagaag aatgtataac tagaataaca aatagagaga agtgcttaaa ggagctgatg 31920
gagctgaaaa ccaaggctcg agaactacgt gaagaatgaa gaagcctcag gagccgatgc 31980
gatcaactgg aagaaagggt atcagcaatg gaagatgaat tgaatgaaat gaagcaagaa 32040
gggaagttta gagaaaaaag aataaaaaga aatgagcaaa gcctccaaga aatatgggac 32100
tatgtgaaaa gaccaaatct atgtctgatt ggtgtacctg aaagtgacag ggagaatgga 32160
accaagttgg aaaacactct gcaggatatt atccaagaga acttccccaa tctagcaagg 32220
caggccaaca tccagattca ggaaatacag agcacaccac aaagatactc ctcgagaaga 32280
gcaactccaa gacacataat tgtcagatac accaaagatg aaatgaagga aaaaatgtta 32340
agggcagcca gagagaaagg tcgggttacc ttcaaaggga agcccatcag actaacagcg 32400
gatctctcag cagaaactct acaagccaga agagagtggg ggccaatatt caacattatt 32460
aaagaaaaga attttcaacc cagaatttca tatccagcca aactaagctt cataagtgaa 32520
ggagaaataa aatactttac agacaaccaa atgctgagag attttgtaac caccaggcct 32580
accctaaaag agctcctgaa ggaagcacta aacatggaaa ggaacaaccg gtaccagccg 32640
ctgcaaaatc atgccaaaat gtaaagacca tcaagactag gaagaaactg catcaactaa 32700
cgagcaaaat aaccagttaa catcataatt acaggatgaa attcacacat aacaatatta 32760
actttaaaca caaatggact aaatgctcca attaaaagac atagactggc aaattggata 32820
aagagtcaag acccatcagt gtactgtatt caggaaaccc atctcacatg cagagacaca 32880
cataggctca aaataaaagg atggaggaag atctaccaag caaatggaaa acaaaaaaag 32940
gaaggggttg caatcctagt ctctgataaa acagacttta aaacaacaaa gatcaaaaga 33000
gacaaagaag gccattacat aatggtaaag ggatcaattc aacaagaaga gctaactatc 33060
ctaaatacat atgcacccaa tacaggagca cccagattca taaagcaagt cctgagtgac 33120
ctacaaagag acttagactc ccacacatta ataatgggag actttaacac cccactgtca 33180
acattagaca gatcaacgag acagaaagtc aacaaggata cccaggaatt gaactcagct 33240
ctgcaccaag catacctaat agacatctac agaactctcc accccaaatc aacagaatat 33300
acatttttt cagcaccaca ccacacctgc tccaaaattg accacatact tggaagtaaa 33360
gctcttctca gcaaatgtaa aagaacagaa attataacaa actatctctc agaccacagt 33420
gcaatcaaac tagaactcag gattaagaat ctcactcaaa accactcaac tacatggaaa 33480
ctgaacaacc tactcctgaa tgactactgg gtacataacg aaatgaaggc agaaataaag 33540
atgttctttg aaaccaacga gaacaaagac acaacatacc agaatctctg ggacacattc 33600

aaagcagtgt gtagagggaa atttatagca ctgaatgccc acaagagaaa gcaggaaaga 33660
tccaaaattg acaccctaac atcacaatta aaagaactag aaaagcaaga gcaaacacat 33720
tcaaaagcta gcagaaggca agaaataact gaaatcagag cagaactgaa ggaaatggag 33780
acacaaaaaa cccttcaaaa aattaatgaa tccaggagct ggtttttga aaggatcaac 33840
aaaattgata aaccgctagc aagactaata aagaaaaaaa gagagaagaa tcaaatagag 33900
gcaataaaaa atgataaagg ggatatcacc accgatccca cagaaataca aactaccatc 33960
agaggatact acaaacacct ctatgcaaat agactagaaa atctagaaga aatggataaa 34020
ttcctcaaca catacactct cccaaactaa accaggaaga agttgaatct ctgaataggc 34080
caataacagg atctgaaatt gtggcaataa tcaatagctt accaacgaaa aagagttcag 34140
gaccagatgg attcacagcc gaattccacc agaggtacaa ggaggaactg gtaccattcc 34200
ttctgaaact attccaatca atagaaaaag agggaatcct ccctatctca ttttatgagg 34260
ccagcatcat cccgatacca aagcctggca gagacacaac caaaaaagag aattttagac 34320
caatatcctt gatgaacatt gatgcaaaaa tcctcagtaa aatactggca aaccgaatcc 34380
agcagcacat caaaaagctt atccaccatg atcaagtgga cttcatccct gggatgcaag 34440
gctggttcaa tatatgcaaa tcaataaatg taatccagca tataaacaga accaaagaca 34500
aaaaccacat gattatctca atagatgcag aaaaggcctt tgacaaaatt caacaacact 34560
tcatgctaaa aactctcaat aaattaggta ttgatgggac gtatttcaaa ataataagag 34620
ctatctatga caaacccaca gccaatatca tactgaatgg gcaaaaactg gaagcattcc 34680
ctttgaaaac tggcacaagg cagggatgcc ctctctcacc actcctattc aatatagtgt 34740
tggaagttct ggccagggca attaggcagg agaaggaaat aaagggtatt caattaggaa 34800
aagaggaagt caaattgtcc ctgtttgcag atgacatgat tgtatatcta gaaaacccca 34860
tcatctcagc cccaaatctc cttaagctga taagcaactt cagcaaagtc tcagaataca 34920
aaatcaatgt gcaaaaatca caagcattct tatacaccaa caacagacaa acagagagcc 34980
aaatcatgag tgaaatccca ttcacagttg cttcaaagag aataaaatac ctaggaatcc 35040
aacttacaag ggatgtgaag gacctcttca aggagaacta caaaccactg ctcaaggaaa 35100
taaaagagga tacaaacaaa tggaagaaca ttccatgctc atgggtagga agaattaata 35160
tcttgaaaat gtccatactg cccaaggtaa tttacagatt caatgccatc cccatcaagc 35220
taccaagggc tttcttcaca gaattggaaa aaactacttt aaagttcata tggaaccgaa 35280
aaagagcccg catcgccaag tcaatcctaa gccaaaagaa caaagctgga ggcatcacac 35340
tacctgactt caaactatac tacaaggcta cagtaaccaa aacagcatgg tactggtagc 35400
aaaacagaga tatagatgaa tggaacagaa cagagccctt agaaataacg ccgcatatct 35460
acaactatct gatctttgac aaacctgagg aaaacaagca atggggaaag gattccctat 35520
ttaataaatg gtgcagggaa aactggctag ccatatgtag aaagctgaaa ctggatccct 35580
tccttacacc ttatacaaaa atcaattcaa gatggattaa agacttaaat gttagaccta 35640
aaaccataaa aaccctagaa gaaaacctag gcattaccat tcaggacata ggcatgggca 35700
aggactttat gtctaaaaca ccaaaagcaa tggcaacaaa agccaaaatt tacaaatggg 35760
atctaattaa actaaagagc ttctgcacag caaaagaaac taccatcaga gtgaacaggc 35820
aacctacaaa atgggagaaa attttcacaa cctactcatc tgacaaaggg ctaatatcca 35880
gaatctacaa tgaactcaaa caaatttaca agaaaaacaa acaaccctgt caaaaagtgg 35940
gtgaaggaca tgaacagaca cttctcaaaa gaagacattt atgcagccaa aagacacatg 36000
aaaaaatgct catcatcact ggccatcaga gaaatgcaaa tcaaaagcac aatgagatac 36060
catctcacac cagctagaat ggcaatcatt aaaaagtcag gaaactacag gtgctggaga 36120
ggatgtggag aaataggaac acttttacac tgttggtggg actgtaaact agttcaacca 36180
ttgtggaagt cagtgtggcg attcctcagg gatctagaac tagaaaaacc atttgaccca 36240
gccatcccat tactgggtat atacccaaag gactataaat catgctgcta taaagacaca 36300
tgcacacgta tgtttattgc ggcattattc acaatagcaa agacttggaa ccaacggaaa 36360
tgtccaacaa tgatagactg gattaagaaa atgtggcaca tatacaccat ggaatactat 36420
gcagccataa aaaatgatga gttcatgtcc tttgtaggga tatggatgaa attggaaatc 36480
atcattctca gtaaactatc gcaagaacaa aaaaccaaac actgcatatc ctcactcata 36540
gatgggaatt gaacaatgag aacacatgga cacaggaagg ggaacatcac actctgggga 36600
cggttgtggg gtgggggggag gggggaggga ttgcattggg agatatacct aatgctagat 36660
gacgagttag tgggtgcagg gcaccagcaa ggcacatgta tacatatgta actaacctgc 36720
acattgtaca catgtaccct aaaacttaaa gtataataat agtaataata ataaattaaa 36780
aaacagaaac attatatcta tctccttgtt ttcatagcaa gtcatcaata catgtttcta 36840
aattgagaaa gaataattac aattgaaggc acatagtgat gaaagaaaaa ttgttgattc 36900
tgacatttgg agctgaaaat atttaatagc tacaatcttt aaaagtcagt atttgaaatg 36960
aaaagttctt ttatatatgg ttgggtgatc tgttggataa tgttttattc ttcttaaagc 37020
aaggtgtcct tgttctttgg acatattttc agtgttgtga gttgtctttc cttgtctctg 37080
tctctttctc tctgttgccc tctctgtatc actgcctttc tctcaattcc tatttcataa 37140
tttggtttag cttatattgt gtcttacaga ctcttcacat gttggcctca gccttacata 37200
ctttcaggac aatctcagta cagacaaggc tgctgcttgt ttttggaaat gactcctctt 37260
gcctgaaatg gctcaagtgc acagcaggtg tatgtgtgtt tgttggtgtg tatgtattgt 37320
gtatgttcat gtatggtgct ggtggaggtg caatcataca acgagaataa tttttgctct 37380

acaatagcta ctaccattta ttttactttt tgcagtaaaa ttttttttctt acccactgct 37440
ctgttcatta acaaacttct agttaggtaa gaaggtgtta ttaactaaga gtttgttaat 37500
gatagtttat cctgttcaga tgtgagaaag aatgttaaaa ctagtgtcta ctctattagg 37560
agatacatta gatgatctca catgcacaat tcctatgtcc attttgcaaa catatttact 37620
gcactgccta ttataatgca tagaattctg ttggtggaaa aatacaagtg tgtccagaca 37680
ttttcgcact ggtttgctat gctaatattg ttatatgttg aaagtgcatg atttgagatt 37740
gactgatgca ttgtcatttt ccccataact tgccatatta gatcactaca ttaatttatt 37800
ttgcttgctg actgtctatt tgttctaaca ttatttttta aaacaatcag aaacttctta 37860
gtagtatatg atactcgcac ataattgttt ggactaaatt agcagtttga ccatccttat 37920
gaggatcaga acaaaatcag atcataagac atgacagaac acatatagct ttgagactaa 37980
aaccttaaaa atgtgactgt tttccttctc ttttttgagg cttcttagag ttgaaagctt 38040
catgtgtttt ctcttttctg tttctgttat actgcctgac agaaacatca ctttacttac 38100
catgcttatt cagtgtacct gcttttcaag gaaaatgtaa tttgtgtgtt ggaaaggaat 38160
ttgaaaagtc actggtcaca tcagcccaca tggatactct ggattcttat ctccatagag 38220
aaggccaata accatttgcc ttgaccacac ccaattccta cttaaaactc tttatctaac 38280
atctaagttt ctcttcttct tctctccttc ttcttatttt tttttttttct tttaaccagg 38340
cttccatgat actggggaca aaaaaagaga agataagaga tgaactagtg gctttctgga 38400
ctctcaactt gaattatttg ttttaaaaaa gtatagggga aaatgggttt gaggaacacc 38460
ttagctttcg atgcttaagc ttttgacttg ttgtaatcct gggtatggaa agaactctga 38520
taactttgtg gctctgtttt ctatttggtt ttgtttttaa tgattctctg gatgttaaac 38580
atctagcaag taatttgatt tgcttaattc tgaggaaaga aaaaatttcc tttgagttaa 38640
atttatgggg taaattcact tacattaaat gtaggaactt taagtgtgca gttcagtaaa 38700
gtttgacaaa tgtgtacagc catgttatca caattccaat tttcaaggct actgtgcagc 38760
tgggaagtgt gagaataggg caagtaaaga tgccacaaag ttcactgtcc ttaccaagat 38820
tcaggaattt tttcttgaat aaatactcac ggaattgttg caagcttttg attactttct 38880
agtatccaga aacagttgac ttttgacttt tttttagtgt tctcatggtt ttcatggagg 38940
gacagttttg ggaaggtcat tactatgtca atcagaaagt tgattcttag atagatagat 39000
agatagatag atgatagatt catcaagatt gtatgaacag tttgttcatt tttattgctg 39060
agagtagtaa cccattatag ggatatacta taattggttc atctatgcat ctgctgatac 39120
atgtttgaga ctctgtatgg atatgtcttc ttttgccttg gataaatact gaagagcaaa 39180
atttctgtgt catagtttta tgatgaacta taaatttata taaaatttta taaagggcca 39240
ccagtttttc acagcccatt gaattatttt cattattata tcttgttatg tgttataaga 39300
gtttcagttg cttcacatta tcattgttaa tctttgtatt gtcaatattt tttattcctc 39360
ccattgaagt gagtgtgaag tgttacaaag ataaacaaag ccagctacca gttaaagtag 39420
taaggacaga ttttattcag taatatacta ttgcaatgga caggtaggtt cactgtaaac 39480
tgaactctaa ttttgttttc acagagataa ctgtgtattc taaaggtaga atgaggaaat 39540
aggaagaggc atgagcaggg ctcaagagag tcagagaagt aaaaaaatta caaaaagtgt 39600
gaaaggagtg ttgatccatg tgaaacccac ccagttttgc caactggctc ctaccctccc 39660
ttgggagcta gtagacaaga gctctttctc cacatattgg ctgaaacaga cagtatcttt 39720
tgttggcagc cttgagtttt ctcaggcagg ttcttccagg agtactacag tcatcccagg 39780
gatatggcct tgagctgtta gccactgtgt ttgtgttttg ttcaagtctt tttaggccaa 39840
ggttgaggcc tagtcaagaa agtgctcaga ggattctttc tagagtttgg tccaggagga 39900
aatctttgtt atgcggtgtt tcataatggg cataatttac attttcttga caaatgcaga 39960
tgttgaattt cttttcatgt gatcgtttgc tgttcatgta tcttctttta tgaagtgtct 40020
gctcacaact tagactttct ttttttttct aattgaactg tttatatttg gaccagtgaa 40080
ttgtgggaga tatttatatg ttcttaatac caatcatttg tcagaattac ttactatgaa 40140
tagtcttcct agtctatggt ttgccttttc attatcttag cattattttg aagaatgcgt 40200
gttttaaaat ttgatgacac acacacactt tatatatata taattatata tatatatttt 40260
attcacaatt agtgctttta ttttcttcct aagaaatctt tgctttcctc aagatcataa 40320
agctattctt ccatttttct ttttttcctg gaggacttgt agatttagca tttacattta 40380
gaatttttat ttccaccaat ttagttttga gtattggcat gtggtgtgtt ataggaagtt 40440
aaatatttta atgaacatat ccatattttc cagtagcagt ctccattaaa ataccttttgc 40500
acctttgttc aaaaatccat tggctacata tacgagtgga tctatttgtg atctctttat 40560
tctgtacact taacctcttc atcatcctta gatccacaca ctgctttgaa tgctatactt 40620
cttagtaact cttaggtgga tgatattaac ttaagggaca gtttttcaca tactttattt 40680
ttttcaagaa agccttcaga aacccctgtt caacttgtgt ttgaagacat taaatgttta 40740
gttaacctga ttttacttca attacaatga tcaattgatt gatgtctgct agaagtgcac 40800
ctaattaatt agttatgtaa atactacaat ggtcagataa ttggacacat ggtactgacc 40860
atggacattt aagtagtgca aattatgcaa caatatttgt tcactttcca aaggggtttt 40920
tgtgttttta tgacagcatt tattttagaa tcagatcatg aagttttgca aacatcatcc 40980
tgctgaaatt tgatagaagt ggctttaaaa ttattgatag atttaggaag aattaacaca 41040
ttgactatat tgaatattct aatgcacaag tgagacgtat gtctttatta tttaggtctt 41100
ctttaatttc gctctgcaga cgtttgtgct tttcagtgtg tatcttttgc acacttttta 41160

ttttgtttc tgagtagctt atgattttat gtgatatttt taagttttat attctactgt 41220
gtattgcttt tatgtaactg tatgagtgtt tttgggtttt tttattttgt tttgttttgt 41280
ttttttagac agagtcccgc tctgtcaccc agattggagt gaagtggtat aatcttggct 41340
ctatgcaatg tctgcctccc aggttcaacg gattctcctg cctctgcctc ccaagtagct 41400
ggggctacag gcacgttgcc atcacaccgg tttcactatg ttggctaggc tagtctcaaa 41460
ctccctacct caaataatct gcccacctca ggctcccaaa gtgctgggat tacagttgtg 41520
agccatcgca cctggcctgt atgagtggtt tttgtgtatt gaatttgtaa ttgtaatcat 41580
atcatcttcc aaaatataca cttaatttcc catttgtata ttttaaaata acttttataa 41640
taaagcatca tgtcatctgt gaataacaca gatttgtttc ttttcaattt gtatactttt 41700
tatttagtag gtaggacttt cagtaccata ttgaaccgac tacaactttt aaataaaaaa 41760
tagccaattg agtaatcaca aacacacaag taaaatctta agggccaata ggatgtatcc 41820
aactgtctga acatggtgtc atttgtatag ctaaatccat taggtgttcc actagcagac 41880
atacatcaat tggtccttcc agcagaagca aagtgacatc gactaaataa actaatgtgt 41940
tcagtaatat agatagataa tataaaactt tctgaaggct ttctatgaaa cataaaataa 42000
gcatgtgtaa aactcatctt tgagttagta taatatacat agctaaatct gtattcccta 42060
accaaatcta taagtaaatt aatgtcacct tctaatgtgt ggacatctga tcatgtgatc 42120
ccatattaaa atttttgaga ctttgctctt cctcatcaaa aataaccaat ccattacttt 42180
ttgtctatcc attcagttat tattgagcca gaatcatttg tagtcagaag agtcattttg 42240
taactcttat gttttattga aatgttattg catattaaca tatgtaagaa aacaaataaa 42300
tgaatataca tatacataca cacatactat agggcatatt gccttcaaat tgaaagctgc 42360
cattctgttt aacaaagagc caaattgtca attagtttta tgcattaagg gatgaggaaa 42420
tgagagacac cagtattata gtcctatatg acccttcatt ctatagaaat atttaagaaa 42480
taccatttta tttttattta tttatttatt tatttatttt ttatttttatt ttattattat 42540
tattatactt taagttttag ggtacatgtg cacaatgtgc aggtttgtta catatgtata 42600
catgtgccat gttggtgtgc tgcactcatt aactcgtcat ttagcattag gtatatctcc 42660
taatgctatc cctccccccc ccccccactc cacaacagtc cctggagcgt gatgttcccc 42720
ttcctatgtc catgtgttct cattgttcaa ttccaaccta tgagtgagaa catgtggtgt 42780
ttggttttct gtccttgcga tagtttgctg agaatttgct gaaaattgag tttccagttt 42840
catccatgtc cctacaaagg acaagaactc ttcatttttt atggctgcat agtattccat 42900
ggtgtatatg tgccacattt tcttaatcca gtctatcgtt gttggacatt taggttggtt 42960
tcaagtcttt gctattgtga atagtgcccc aataaacaca cgtgtgcatg tgtctttata 43020
gcagcatgat ttataatcct ttgggtatat actcagtaat gggatggctg ggtcaaaaga 43080
aataccattt tacttttaaa ttagtgcaag gaatcaaaac catacagata aaaactgaga 43140
taaggcttta aactcaattt gaggaaaaaa atagctatgc tcaaatctct caaatttcat 43200
gggatgtaac aaatctacat tttgattttg atatatagct atattgtgtt gactgatttc 43260
tcaaagaata accaccattt gtgatctaag agaattaggg atgcctcttc tgtcttggtc 43320
agagcccatt ttgattcatt tgtgctaacc tggaactatc agtgcactgg aaaatacaga 43380
ctaatgaaca catgtctagc ttcccttatc atctaacaaa tgcagtgtaa acaagtttta 43440
aatattcttc cttctatttt cagccctatt cctaagtcaa caattttatg actcatcaat 43500
tttataaacg gtcctggtct cagacctgtc tttgattctt agcttgttaa tgactttcaa 43560
tagggcagac tagttccagc ataaatccag aacgtccgtc ctccaaagga cctgggccct 43620
gaccaatact tctgcttcct tctactgatt atgttgtcac tctgctgctc tccacatcct 43680
tacaaatcct tcctttcctc aagtctttca gctgccacac ctttctaccc ctcaccctca 43740
gtaaaggacc tcaccccagt gtctatagac aaaactggag agttctgatg aaaaccttct 43800
tggttttatt atattaaaat tacaaaactt atctatctat attttccttc ttccatatta 43860
ttgaaagtga tggctagttt ctctgtttgt gctaaagctc tgatgaacac atgtaacaaa 43920
actgtacata aatgttgtat agggatttca tcttcaaagg agaatctagc tgttcaaact 43980
atcaaataaa tgttaagccc ttgaggaaca aatgatatta taattggttc tttagaatta 44040
ctttaaaaat actctaggga atatagctaa gaaatatata tcttgccacc ttgctttaca 44100
aactatggtg agttaacatt aacaggtagg tcatttgcta catttgttca ctaaaaatac 44160
aacttttttct gtgtgtggag attagtatcc aatcaataga ggaaacccat attttcatca 44220
atttcactgg acttttctat tcatatgtta acatttcatt gactctaaaa tcaattatat 44280
gcaagctcct ctgatgtata tgtgagagct aaataagcaa tttctttagg gtgcaaattt 44340
ttgagcaata atttttacag tcatagctta ttttcatgaa gaactcagag agagtggctt 44400
gagttggcag caaagactgc ttttaacaat gttctgtgat tctctaacaa tattttaata 44460
agtagtttag tggcatataa atggagacaa aaaaacttca ctctcttcct tcaggtgcta 44520
taattgttta tttactgttc aaatgttgca ttctcagaaa aggctacttt gaatagtcat 44580
ctaaaagaat ctcacattca ttgttcttca cagacccagc ttttacttga aattataata 44640
tgtatctttc tctgtgggtt taaccattat ttctcatatt ataatgtgag ctccatgaaa 44700
ctggaaaatt tgttcactcc aataacccca ggaattgtaa tacacacaaa aaaatcaagc 44760
tatgcagtca gtgtacatga ataaacaggc aggtaaaatt tcaaagtgaa aaaacagttc 44820
ttggttttgt atattcttag ccatatatca agcactaagt ttaactaatt taatgttcac 44880
tgacaatcat aggtttctgt ttgtttgttt gtttgtttgt ttgagacgga gtctcgctct 44940

gtcgcccaga ctggagtgcc ctggtgtgat ctcggctcac tgcaagctcc gcctcccggg 45000
ttcacgccat tctcctgcct cagcctcccg gtagcttgga ctacaggcac ccgccactac 45060
gcctggctac ttttttgtat tttgagtgga gatggggttt cacagtgtta gccaggatgg 45120
tcttgatctc ctgacctcgt gatctgcccg cctcggcctc ccaaagtgct aggattacag 45180
gcgtaagcca ccgcgcccgg ccgacagtca taggttttaa gtgaaaatgc taccttaatt 45240
ttttttcttct gttttgtatc tgaaataata agatatcaga catggaggga aactgaagat 45300
tggacttcct acatttattt atttaaatga tgctaatatt aatatctgat attcatttct 45360
ttgttttttt tgttccctca caaaccttga aaataattac caaatgtaga attcaccaag 45420
aagacaagta taaatgcatt aatgtataaa tttataaatg tatatatata agagacagaa 45480
atagagaatg tgcctgtgtg tgtgtgggtt tgagatacaa aaagagacag aaaacacaca 45540
catataaaga gaaaacgttg tgcagaaatt ctgatgtgtt tattgaaaaa agtaattggt 45600
gcctagatga atagatgaat gagtgaatga taaatggatg aaacaaatgc caaatctgga 45660
tcagagagaa tcctcacatt ctttgtcact ttcagttttc aagaaataag aagatgttgt 45720
ccccaaacca caccatagtg acagaattca ttctcttagg actgacagac gacccagtgc 45780
tagagaagat cctgtttggg gtgttcctgg cgatctacct aatcacactg gcaggcaacc 45840
tgtgcatgat cctgctgatc aggaccaatt cccaactgca aacacccatg tatttcttcc 45900
ttggtcacct ctccttttta gacatttgct attcttccaa tgttactcca aatatgctgc 45960
acaatttcct ctcagaacag aagaccatct cctacgctgg atgcttcaca cagtgtcttc 46020
tcttcatcgc cctagtgatc actgagtttt acttccttgc ttcaatggca ttggatcgct 46080
atgtagccat ttgcagccct ttacattaca gttccaggat gtccaagaac atttgcatct 46140
ctctggtcac tgtgccttac atgtatggct tccttaatgg gctctctcag acactgctga 46200
cctttcactt atccttctgt ggctcccttg aaatcaatca tttctactgc gctgatcctc 46260
ctcttatcat gctggcctgc tctgacaccc gtgtcaaaaa gatggcaatg tttgtagttg 46320
caggctttac tctctcaagc tctctcttca tcattcttct gtcctatctt ttcatttttg 46380
cagcgatctt caggatccgt tctgctgaag gcaggcacaa agccttttct acgtgtgctt 46440
cccacctgac aatagtcact ttgttttatg gaaccctctt ctgcatgtac gtaaggcctc 46500
catcagagaa gtctgtagag gagtccaaaa taattgcagt cttttatact tttttgagcc 46560
caatgctgaa cccattgatc tatagcctac ggaacagaga tgtaatcctt gccatacaac 46620
aaatgattag gggaaaatcc ttttgtaaaa ttgcagttta ggcctgtgtt tatttgtaat 46680
ccctaagtgc ctgtggggta acaaactgaa atggaaaaac ctagtgtagt tattatttaa 46740
cagtatgggc tcttagtaac cactttagtt tcttctcaaa attaacactt tgaagattta 46800
gtttttaaaa ataaaaagct taatgttgaa attaataatg ttttatttgt cagagattct 46860
atgaatataa attgtttggt ttctaataat tctgtatgaa aatactgggc ttacttactt 46920
gggggtaaac ttttaactaa cttagcctga gagaaactct gaaaatccaa ccccttttgaa 46980
tattactatc catagagcca tagtgaggct ccagataggg tggctaattt tacattttta 47040
aaaaaatgat gccgggcatg gtggctcaca tctgtaatcc cagcactttg ggagccagag 47100
ataggtggat cacctgaggt caggagtttg agaccagctt ggccaacatg gtgaaacccc 47160
gtctctccta aaattacaaa attagctggt tgtggtggca catgcctgta atcccagcta 47220
ctcaggaggc tgaggctgga gaatagcttg aacctgggag gcggaggttg cagtgagtca 47280
agatcacgcc attgtactta ctccacttgg gcgacaagag cgaaactcca tctccaaaaa 47340
aataaataaa taaataaata aaaagaaaac gaaaaaagaa aaaaattgtt tctacttcaa 47400
tgtggtgtct tgaacaatag ttctgaaacc tgtgaactta ttgtaatcac tgtgcaaaaa 47460
atgtaaatat atgaatcacc tataagttaa attttctgta agtattctca aaacccacac 47520
tgtaaacatc gtgagaagct tcttattggt tcatgtggta ataccagtta tggtcctcaa 47580
ttctgaaaat caactctcaa caccctttt ttttttttga gttttcaagt gcatggatcc 47640
atttctatca tcaaaatact tcaattgcat tccacgtatg ttttccagtg aagtacctct 47700
gacttaggaa taatttagta ggaatcaata gggtttatca attgatattt gcatatgcat 47760
aaatgacttt gttagtggaa aaaggggtct tgatccagat cccagcaggg ggttcttgga 47820
tctacagagg aaaaatttca agataagctg caaagtgcag tgagaagaga tagttcatta 47880
aatgctactc caaagagatc atttattaaa ggctattgca ttacagagta aggtgttccc 47940
agaaagcaag tggaggaatg caccctcttt aagttttttt aatggtcttt tatctacgta 48000
cagactaaag taagttgtgg ctacatgtgg gtaggctgac ggcatgacaa aatttatcat 48060
tgtgttgatt taaagaaaac gatccttgat attttagtgt gtgcataact attattatca 48120
tgataaaagc atatactctt atgagaattg ggacacctag gttctcttgc tgcattatta 48180
gcgtgttctt gtaggtattt ttaggctgtt tcctaaacta taagcatctt atgaacatgg 48240
gtcatgaccg gccaggactg tgccttgtta acctcaagac agagttgatt ttaaaatgtt 48300
gtcaccctgg atctcctaga ctcctgcttc tttaacaccc tgaaacagta taaaaattaa 48360
taggggcgta cgtatctttt tggaggaaat agtgtttatt ttgcttcaca caactgtctg 48420
tggcacatta tttttttaca ttagggtatt ctagttctca aagctactac agggaaaatc 48480
aatcacaaag aggctaaata aagatacaca ttaatcaatg tattatttag tcaaccagat 48540
gactgcatat taatacatac atatttagcc ctattaaatt gcctcttatg tttgacattt 48600
taatattttg cactctgctt ttaactaaca ctatacagaa aatgctacag aaaatgctta 48660
tatgatatca actactgtat ctactataaa atgttcaagt ggttttccag aatattaaac 48720

agatgactaa cagtttttgt aaaattgggt tatattctta agtgtcagtt ggttgaataa 48780
gtctggttct atccaaatct gcttttagaa acagtaactg aaaaagaaag aaggcaattt 48840
tgggcagcaa caatttaata tttacctttt agcactattt gtgtatttgc ctggttttgg 48900
caatggcagt ggcacaacag aagcaggtga tctgtagaga caagagtgaa gcagaagagc 48960
tgggaactga gtgacacagc gtccttactt ccttttagag gcccagcatc atttcatctg 49020
tgggatattt ctttatctta tgacaacatc atttaatctg tgggatagct ctttatatta 49080
ttctttttatc ttcctctttt ggttaggcaa agcattgtgt tcaatttcta gcaagtagag 49140
tggtctgata ctctatttga aaactgtatt ttagaattgg tacaaacaca ctctatattt 49200
tgcctaaata actgtatact tccttttatt ttctagttat gagaattgct gtggcttgca 49260
gcaaccaaac ttaaactttg acacctttgt aacccaaatg cttgacctgt ttgtaatctc 49320
tttttttaa aatatgtatt tatttattct atttgtaatt taatttttta attgtgcaag 49380
cccttgggta aaagttatct cagaaattct ttggagatag aatagaacat aaatggacat 49440
gaatagaaat agtgacaatt gatgtaatca tagtgtattg ctaattttag gattactgtt 49500
ttttctttac ttaaatgaaa atgaagtagt gagagaatta ttatcttcat ttacaatatt 49560
tggagtctta gacacagaga gatgaaatga cttgcctaac gtcatttggc cactcagtag 49620
cacaaaagca gatttcttgg ttttaaattt tctctctttc taattcatac tgaagtggtt 49680
attaagccat tcttcaacac acttcaaatg ctgggggggaa tattggagac tatgaattaa 49740
taagagtgtt actatagatt aaatagtggt tattatgtga tagtgagatg tatcctaaca 49800
acagctaatg tgaatttttcc ctgacagtca ggctaatgtg atcatttata tgcatgataa 49860
ttaaaattgt cttaatatcc atgtgaaggg agaaaatttt gatatatggg aaacctttat 49920
gcttcttagg ctgcatttta ttaatatatt ttcccaatta taaaagcaca catatttata 49980
ttaattaatt tccaggtaca gacaaatata aaaacataaa atttattaag gacctgctat 50040
gacaaggaat aagttttgtt tatattttgt tatatttctt ctcaataaat gtatatgact 50100
acatgttact ataatgcaca aaacaaatat aaagttactc accatataat aatatgaata 50160
cagattttca tttgatacta taatgtatgc acctttacat agattaaata ttttatgaac 50220
atttttaatg tcatatttta attagcttgt gtgatcctaa tacaatatta atacatgtca 50280
aatgtagaaa accaatgaac catattggca tacataagaa atagaagcaa tgcttacttg 50340
taatcaacta cctgaaaaga gttgctctca gcattcacat gtgttttatt ccatttgatt 50400
atcttgcaaa tcatgattcc atcatatcaa tgtatcttag tattttgatg accgacgtat 50460
ctgtacgtct ctctctaatt gttccttttg ttatgttccc agaagtgtaa ttgctgaatt 50520
aaatgttttg aacattttta aaggtcttgc taaacatttc tatggataaa tttttatcca 50580
gaaagcttgt aacaaacctg taccactaat gtttgaaaac aatcatttta ctataactct 50640
atcagcagta ggagctcttg tttatggtta tatgctactt tgatctctca aaggaagttt 50700
tagaaacaaa ttctgtttta atttgcacag attagaaatt gaagaatata tttaacaaat 50760
aaaggctaca tataataatt ttctttcatg ttatgatggt gtcttacatg aagaatattt 50820
gtatttttat ggagaaaaat tatattttgt ttattgttgt tttgtaactt tttcttatgc 50880
cacaattgaa taaacaagca cttatatggg gttattggct ttcattgcaa tattttatgg 50940
gaatatcatg aattatcctg gatataattt tgttgcatgt tttaatgttg taaattactt 51000
ttgtgccctc acatttttaa ccaacttcct ccaagtgtta ttaagcactg ctttcccact 51060
gatttgtgat gtcaccttta ttatacctta agtcaatcta tttctgccca tttccactgt 51120
ttttttgtc tgaattataa atatatttta aactaaattc ttatcattaa attgaacata 51180
tattttaatt tctggtgaag atatttttgt cttcaatttt ctcttaggag attgaaaatc 51240
agatactttt agctatccaa atttttaact cagaaattat aatattatga ttcaatttag 51300
aaactgtatg tatagttgta taattttgaa tgtagttgta catacaatac ttatacatta 51360
taaagcaatt aaaaaataag gaaaattatt ttactgtcta tggtggccat agggagaact 51420
gcttaaggat gaaatatttt gaatgagttg tttgttttcc agaaagtcag gaagatgctg 51480
gtgtctggta agtgaggaga aaacacctaa ccttggtata ttaattctat ggatttatta 51540
acactgttat acaaatactt atgttttcac cataccagtg tatggtgaaa tcatcatgta 51600
acatttttt tcaacaaata tctctaaata cagtctcttt ttcttacaaa gaggtttttt 51660
tcccagcaat taccccaaa agaactgaga gggccacata tgtttttcca ctatgagtca 51720
cttcaaacac tgaacaaaaa aggattaatc agaaagaaaa taaatattta aatttaaata 51780
atcagagaat aatatatctg tgcttcataa cagtgggaaa aatttgtagt aaaaagagta 51840
agcaaattgc tttaatttga agataaaact atttaaaatt aagaggcacc ctacaatttg 51900
gcaatgtatt tgccacaaac aagataaatt ggtaaaatta ttgaaataca agaaattaat 51960
aaaaataatt tgaagcacta aagaatcaat atataaatat caaagactat aagtaaaaaa 52020
ttgattaaaa tacaaaaatg ctaaaattaa cacacaaaat attagcaacc aaaaatatga 52080
tgtaagatat tgtttgatat atctacattt aaaagagctg atatagattg gcactttcat 52140
acattgattg attaagggta taaagttttt ctgaaaataa atttaatatt taataggttg 52200
gtacaaaagt aatcttgttt ttttgccatt aaaagtaaca gcacttattt tttttcttac 52260
caacctaata ttattatcat attaaggaga caaaattttc ttatgctttt attcagctat 52320
ttcaatttta agaatctact gaaggcataa gaaacttgtg tgaagagttt catggtgttg 52380
tttttatttt ataatcaaga caaagtaaac acatgatgta ttaaatacaa atagttcctg 52440
cttggcatgg taccatgtaa cagaaacctt tccatactgg aaacgtgtct tagtttttgt 52500

ggttcttagt tcctacaaaa tcacaaagtg agaactgcct gcatttaata actggtaaaa 52560
tacaggcatc caccagatgg gatagaagct gtctctaact aagtagaatg tcattagcct 52620
tgatcatgaa ggaaagtgta aagtacaggt gaaacgtttt ttgtaataga aagttataca 52680
gccattaaag tgaggatgtt cctagttaag aaaaaggtaa ttcattttat gtaatatatt 52740
atgtgcaaag atatatccta tgctgtatac agtgtgatat cttgtaaaat attagggata 52800
aggggaaata aaacttgtta ttaatagcta tgtccatatt taaatagtca ttgctattgt 52860
gagaaaataa ggcaacaaaa actgtgtatt ttcagtttag tccagatctc aggatgtata 52920
attacaggat acttatttat taatattgga atttaaatat gaaagtaagg aaaagcaagg 52980
aaaaagcaag gtaagaaaaa gaagaggtcc ccaattctgg tttccagcac agaatgacca 53040
atgaacaact tctcagagaa aaatgcccat gtgaaaattc cgtaacccat gggtgaggtg 53100
gggcacactc ttggagcata aaaactggaa aaaaccacat tagaattgtt aagaggagca 53160
gtttcacttg cttctttccc agccaaacac agtgtcacac tgaaagagat tccctgggcc 53220
tgtggtttcc ccagtgaaga aaagagagcc cacggcagac atccagcttc cctatgttcc 53280
agagaacttc ccaagaggcc cattctgtct aaccttgtgg gtaataacgg aagacttggc 53340
aaagcttggc caccccagtc agctcataac aaagttaagg ggtgtagctt acggcaacca 53400
gcacccgatc ttggtggtgg catcatgttc tagctggcag ctttatgtga ccagagagcc 53460
cagccactag cattgctcac atgtggagct gagttggcag gccctttcgg tcaggaggca 53520
tgttgaacag ttctgtctgg ctgtggtacc agtcagtaag ctggaccagc tgcagagcac 53580
agcctagagc accacccaac cacaaagcac aacttgcagc ccaccccact caggtagaga 53640
tccaagccag caaccccact caaccgctgg gcatagcctg cagtcccatc tgacgacgga 53700
gtctggtcaa ttatctctcc aaactacaga gcacatccag taaaacttcc caattatgga 53760
tcagcttgca gccccaaaaa actgcaggaa atagccagtg gccctatctg gccaagagat 53820
ctggtcagtg atctcaccta aatttggagc aaaggcagtg accccatcaa tctgtggacc 53880
acagccagca gccccactcg aatacatacc acagtgaatg gtcttacctg attatggagc 53940
ccagtcagca gccctgagtg attttggatc ctagccagca tactcaccca cctcagagca 54000
caggcagcag tcagaggtca tgtgactaga gtcagctttc aaattcatca agtcctggtt 54060
ccacctctta ctccctatga aatctcaggc aaattattta acctctactc ctatgcccca 54120
atttgtttgt ttgtttgttt tcacagtact tatcatcacc tgccatatat ttgtttgtgg 54180
gtgtatttgt catctatctc ctccagcaaa aaacattatc aataaagtag tttttaattt 54240
tttttgagca acagctatat tgagttatat agtggctata ctaatttaca tttttgccaa 54300
taatgtacga gcattccttt ttctgtgtat ccttgcaaac ctctgttcat ttttttgtct 54360
ttttaatgac aatcattaca aattggttga ggttatactt cattatggtc ttgatttgca 54420
tttcatttat gattagtgaa ttaaaccttt ttttcatata cctgctgaca atttgaatat 54480
cttctttttg aagcttttga taaaatccaa catctcttca taataataat aataaaaaaa 54540
tactcaacaa actaggcatt gaagagacat acctcaaaat aataagagcc atctatggca 54600
aacccacagc caacatcata ctaaatggga aaaagctgaa agcattccac ctaataaatg 54660
gaacaagaca aagatgtcct ctctcatcat tcctatttca catagtactg caagttctag 54720
ccagagcaat aaagccagag aaagaaatat aaggcatcca aataggaaaa gaagtcaagc 54780
tatctctctt cactgataat atgattttat acctagaaaa ccctaaagac tacaccaaaa 54840
agctcctaga tctgataaac aactttatta aagtttcagg atacaaaatc aatgtataaa 54900
aatattagca tctctatact cccataacat tcaagctgag agctaaatca agaatgtaat 54960
cccatttaca atagccacac gcacaaaaat aaaacaccta ggaaaacaac aaaccaagga 55020
ggtgaaagat ctctacaagg agaactacaa aacactactg gaagaaatca tagatgatat 55080
aacaaatgga aaatcatccc atgtttatga attagaagaa tcagaatgct aaaacagcta 55140
tattgcctaa agcaatctac agattcaaca ctattgccat caagctaatg acactattta 55200
catagaattt gagaaaactt gcctaaaatt tatatggagc caaaaaagag actgaatagt 55260
caaagcaatc ctaagcaaaa agaacaatac tggcagcatc aaataaccca acttcgaact 55320
atactacaag gctacagtaa ccaaaaaagc atggtactgg tacaaagaaa gacccataga 55380
caaaatgaaa cagtgaaccc agaaataaag tcagacatct acaactgtct gatctttaat 55440
aagttgatag taataaacaa tagggaaaga actacctatt caataaatgg agccaggata 55500
actatttagc catgtgcaga agagtgaaac tggacccata cctgtcacca catataaaaa 55560
ttaactcaag atggattaaa gacttaaata taaggcctaa agctataaaa atcctggaag 55620
aaaactgagg aaataccact ccggacattg gcctcagcaa aaaatttatg aataagtctc 55680
cagaagcaat tgcaataaaa ataaaaattg aaaaatggga tgtaattaaa ctaaagagct 55740
actgcataga aaataaatta acagagtaag caatctaaag aatgaaagaa aatgttcaaa 55800
aatgataaat ccaacaaaaa tctaatatcc agaatatata agaaacataa acaattcaac 55860
aggcaagaaa cctcccccta aaaccattaa aacacatgga cacagggagg ggaacatcac 55920
acactggggc cttttgtggg gtgggaggct aggtaaggga taccattagg agaaatacct 55980
aatgtagata acggattgat gggtgcagca agccaccatg gcatgtgtat acctgtgtaa 56040
caaacctgca cgttctgcac atgtacccca gatcttaaag tataataaaa aataggcaaa 56100
agacataaac agacacttct caaaatgaag catacatgtg gccaaaaaat atattaaaaa 56160
tacagtatca ttaatcatca gagaaatgta tatcacaact acaatgagat aacatctcac 56220
accggtcaga atggctatta ttaaaaagtg aaataaataa cagacattgg tgaggttgtg 56280

gagaaaatgg aatccttata cacttttggt gggaatgtaa attagttcat ccactgtgga 56340
agcagtttgg aaatttctca aagaacttaa aacacaagta ccatttgacc cagcaatcct 56400
attactgggt ataaacccaa aggaaagtaa ataattttac caaaaccaca tgcacttgtg 56460
caatcattgc agcactatac actgtagcaa agacatggaa tcaacgtaga taaccatcaa 56520
tgatgaacgg gataaagaaa atatgataca tacacactat tgaatactat gaagccatga 56580
aaatgaaatc atgtcttttg cagcaacata gatgcagctg aaggccatta tcctaagtga 56640
attaatgtag gaacagaaaa ccaaatgtgc ccatgttctt attaaaggta tgagagaaac 56700
attgaataca catgaaaaat aaaggagaac aagcaacact gggggttaat agaaggggaa 56760
gagctgggga gtgcagactg aaaaatgatc agaccaatgg aacagaatag agagcccaga 56820
aataagacca cacatctatg accatctgat cttcaacaaa gctgacaaaa acaggcaatg 56880
ggaaaaggac tccctattta ataaatggtg ctgggataac tggctagcca tatgcggaag 56940
attgaagctg gacccccttc cctataccac acacaacaat caactcaaga tggattagag 57000
acttaaatgc aaaaccccaa actgtgaaaa ccctaggaga aaatgtcggc agtatcatcc 57060
tagaaaaaga aatgggcaat gatttcatga caaagacaca aaaagcaatt gcaaaaagaa 57120
aaaagcaaaa attgacaagt aggatctaat taaacttaag agcttcttca cagcaaaaca 57180
aactgttaac agagtaaaaa gacagcctat ggaatagaag aaaatatttc caaactatgt 57240
atctgacaaa ggtctaatat tcagcatgtt taagaaactt aaatttacaa gagaaaagca 57300
aacaacacat ttaaaagtgg gctaagatgc tggcaacagt ggctcatgcc agtaatccca 57360
gcactttggg aggcagaggc aggtggatta tctgaggtca ggagttcaag accagcctgg 57420
ccaacatggc aaaaccctgt ttctactaaa aatacaaaaa ataactgggc atggtggtgc 57480
gtgcctgtaa tcccagctac tctagaggct gaggctggag aatcacttga actcaggagg 57540
tagaggttgc agtgagccga gattgcacca ctgcactcca gcctgggcga cagagtgaga 57600
ttctgtctca taaataaata aataaataaa taaataaaat acataataaa agtgggcaaa 57660
gaacataaac acttctcaaa agaagacata catgcagcca acaattatat gaaaaaaggt 57720
caatatcact gattattaga gaaatgcaaa tcaaaaccac aatgagatgc catcttacac 57780
cagtcagaat gactattatt aaaaagtaaa aaaataacag atgctggtga ggttgcaaaa 57840
aaaaaggaac atttttatac tgttagtggg agtgtaaact cgttcagcca ctgtggaaag 57900
cagtatggca attcctcaaa gagctaaaag cagaactacc atttgatcca acaatcccat 57960
tactaggtat atacccacag aattataaat cattctacca taaagataca ttcatgtgac 58020
tgttcatttt ggtactattc acaatagcaa aggcatggaa tcaacttaaa tgcccatcaa 58080
tgacagtttg gataaagaaa atgtggtgca tacacaccat gtaatactac atagccataa 58140
agaagaagaa gatcgtgtct tttgcaggaa catggatgga tctggaggct attatcttta 58200
acaaactaat gctgaagcag aaaaccaagt accgtatgtt cttatttata agtgggagct 58260
aaatgattag aacttacaaa cacaaagaag gaaacaacag acactggggt cttcttgagg 58320
ggggagggtg ggaggaggga taggagcaga aaagataatt atgggtacaa ggcttaatac 58380
tctggtgatg aaataatctg tgcaacaaac tcccatcaca catgttcacc tatgtaacaa 58440
accttcacat gtatccccaa aactaaaata aaagtttttt tttttaaagg aaaactgttg 58500
ggtactatgc tcactacctg ggtgatggga tcatttgtat accaaatatc agtgacattc 58560
aatttaccta tgtaacaaac ttgcacatat acccctgaaa ctaaaaatag aaaaaattat 58620
taaaatacaa caaaaaagaa aaaagatgaa aaagagatat ctataatcat ttgcacattt 58680
caatcagatt tttaaaaatt tatttatttt gaaattgagt tgttcaaatt tcctatgtaa 58740
cctgaatatt agtcccttgt cagatgaatg atttgtaagt attttctccc attgtgtagg 58800
ttgtcttttc actctataga ttacttcttt tgctgtgcag agatttcttt tgtttgatga 58860
aatcctgttt atctatattt ggttttgttg cctgtgcttt agaggcctta tccaaaaact 58920
ctttacctag accaatgtct aaaagtattt tcccagagtt ttcttctagt agatttatag 58980
tttcaggtca tacatttgtc ttaaaacat tttgattaga tttttgtata gtagtctatg 59040
atagtctttt atatgtctga ggaattgatt gtaatgtcac ctttgtcatt tctgatcgtg 59100
ctgatttgga tcttctttct tattttcttt gttaatctag ctagcagtct atccattttg 59160
tgtatccttt caaataaact tttggtttta ttggtttttg gatgtatttt ggggtctcag 59220
tttcattcag ttttactctg atttagttat ttatttttctt ttactggctt tggagttagt 59280
ttgttgtttt atagttcctc taggtgagat gttagatcat taatttgaga actttctaac 59340
tttttgaggg aagtgtttag cgctgtacac tttccttta acactgattt tgctgcatcc 59400
cagagatttc ggtaagttgt gtctctgttt ttatttattt caaatatata ttttgttaat 59460
ttctggttta gttgtttgcc caaaagtcat tcagtagcaa attgtttaat tcccatataa 59520
ttttgtggtt ttgagagatc ttggtattga tttttatttt tattccagtg ttctccaaga 59580
gggagtttga catgatttta tttcttaatt gattgacact ggctttatgg cctagtatgt 59640
gatcaacctt ccagtatgtt ccaggtgcag aggacaagaa cgtatattct gtggttatgg 59700
gtggagtact ccgcagatgt ctattaggtc caattaatca agtgtcaaat ttaagtccag 59760
aatttattag ttttcagacc ttgatgacct aaggctgtaa ttggggtgtt gaattcccct 59820
actgctattg tgtggctttt taggtctttt tgtcagtcta gaagtacttg ttttatgaat 59880
ttgggtgctc caatgtgggg tgtgtatatg tgtaggatag ttaagtcttc ttgtttaatt 59940
gaaccctatg tcattaggca gtgccctttt tgtcattttt tccagttttt ggtttacagc 60000
ttcttttatc tgacataaca atagcaaccc ctgctgtctt ttatttttca tttgtctagt 60060

agattatttt tcatcctttt actttgagcc tatgtgagat agatctcttg aaaattgcag 60120
atggatggat cttatttttt atctaacttg ctactctagt atatgcctgt ctttataggt 60180
aaatttcttg aagacagcac atacttgagt attgattttt gtttgtttgt ttttcttttg 60240
ttttactttt ttttccattg agcagtcact cagcgtcttt taattgtaga attttgtcta 60300
ttaacattga agagtattat tgttaggtaa gagtttacta ctgccatttt gttatttctt 60360
tgcttgttgt tttctaaatc ctttatttct tgtgttctat cttactgtat tactctgtag 60420
ttaagtaatt ttctttagaa gtatgttttg attccatgct atgtttttgt gtatctgtca 60480
taggtttttg ttttgtgggt accgtgaggt ttacaaaaaa aatctatagt ataacagttc 60540
gttttgaaca gataatgact taatttttgat tgcaaataaa agcatcaaaa accaactcta 60600
cacttttaat atgccctaca cacattttga cttttggtgt ttcaatgtac atgtttttat 60660
attgcctact tctaaactgt agttattgat ggttttaata gttttgtatt ttattcttca 60720
tacttatgat acaagtggtt tatttaccac aattatagta tgacagtatt ctgattttga 60780
atgttctttt acgagggaat tttatacatt ccaatgtttt cctgttacat gttagtaccc 60840
atttctttca gactgaagaa ctctctttag catttcttga aaaacaggtc tggtgttgat 60900
gaatttcctc agcttttgtt tttctgcgaa aaaaaaatct ttatttcacc ttcatgttcg 60960
aagaataact ttaatgcaca taatattctt ccttgaaaga ttttttctct tatcatcttg 61020
agtatagcat accatttcct cttgacctgc tagcttctgt gagaaatctg ctgaaagcca 61080
tattcaagct ctgtattact ctgttctcac attgctaaac aaaattacct tagcctgagt 61140
agtttataaa gtgaagaggt ttaattgaca cacagttcca caggctgtac aggaggcatg 61200
gctgggaggc ttcaggaaac ttacaatcat gatggaaggt gaaggggaag caagcacatc 61260
ttcacatggc gacagaggag agagggtaaa ggggaaagtg ctacacactt ttaaacaaac 61320
agatctcatg agaactctat cataagatag cactagggga atggtattaa accattagaa 61380
aacaccccga tgatccattc acctcccacc aggccccacc tccaaaactg ggggattaca 61440
attcaatata aggcttaggt gggaacacag agtcaaacca tatcattctg tcccggtccc 61500
tcccaaattt catgttcttc tcacatttca aaacgcaatc atgtcttccc aatagttccc 61560
caaagtttta actcattcca gcacaaactt aaatgtccaa ttcctcatgt ccaaatgtcc 61620
aaatgtccaa aatctcatct gagacaaggt ccacaggtaa gcctgtaaaa ctaaaaacaa 61680
gttatttatt tctaagatac aagggagtta caggcagtgg ataaatgctt ctgttccaaa 61740
aggaagaaat tggccaaaat aaaggagcta catatgccat gcaagtccaa aatccagaag 61800
ggcagtcatt aaatcttaaa gctccaaaat aatctccttt gactccatgt ttcacatcca 61860
ggccacacta atgcaagggt tgctctctca aggcctgggg caactccacc ctcatggctc 61920
tgcagggtgc agccctgttg gctgctttca caggctgatt ttgagtgctt gtggcttttc 61980
cagatgcatg gcacaagctg tgagtggagc taccattctg gtttctgaaa gatagtgtcc 62040
cttttcttac agatccatta cgcaggaccc agtgggaact cggtgtgggg gctccaaccc 62100
cacattttcc ctctgcattg cactagtaga ggctctctgt gaggattctg cccctgcagc 62160
agacttctgc ctggatatcc aggcatttgc atacaacctc taaaacttag gcaaaagttc 62220
ccaagcctta attcatgtct tctgcacacc cacaggccag acactatgtg gaacttgcca 62280
aagcttgggg cttgcaccct ctgaagcaat ggcccaagct gtaccttggc acttttagt 62340
cctgactgga gttggagtgg ctaggacgca ggtcatcatg ttctgaggct acacagagca 62400
ctgcagacct gggtctggcc cccgacatca ttttttccta tttgccctct gagcctgtga 62460
tgggagtggc tgtaataaag gtctctgaaa tttcttggag gcattttccc cactgtcttg 62520
gctattaaca tgtagctcct ctttatttat gcaaacttct gcagccttga attcctccct 62580
ataaaatggg tttttctttt ctaccacatg ttcaggctgc aaatttttca aactttcatg 62640
ctctgctttc cttttaaata taaattccaa tttcagacca tctctttgaa aatagagaag 62700
aagccaggct aattcttgaa tgctttgctg cttagaaatt tcttccacca gatacctgaa 62760
ttcatctctc tcaaattcaa agttccacag ctctctagag caggggcaca atgccaccag 62820
tctctttgtt aaagcatagc aagagttacc tttactccac ttcctaataa gttctacatc 62880
tccatctgag accacctcag cctggacttc actctccata ttactattag tattttgttc 62940
accaccactc aacaagtctc taggaagttc caaactttct cctcttcctg tcttcttctg 63000
agccctccaa actgttccaa cctctgcttg ttacccaatt acaaagtcga ttccacattt 63060
ttgggtatct ataggaatgc cccacttctc tggtaccaat tttatgtatt agtccattct 63120
cacatcactg taaagaacta actgagattg agtaatttat gaagaaaaga ggtttaatcg 63180
attcaccgtt ccttaggctg tataggtagc atggttgaga aggcctcaga aaactaacaa 63240
tcatggtgga agggcaaagg gaaaacaagc acatggcaag aggagaaaaa cagagcgaag 63300
ggggaagtgc tacacacttt taaataacag atattgtgag aactctaaca caaggcaaca 63360
ctaggggaat ggtgctaaac cattagaaat caccccatg atccaatcac ttcccaccag 63420
gtcccacctc cagcagtagg agacggcaaa tttagatgaa atttgggtgg agacacagag 63480
ccaaaccata tcagtctcca ttaaatgtga tatttttctc ttgctgattt gaatattatt 63540
tctttgtctt tgatttttcc taatttgatt acagtgtaac ctgagatatt actgtttgta 63600
ttgaatttga ttggtgagca atggtctttt ctacctagat gctattgtgt ttctccagaa 63660
ctgagaaatt ttcagttatt atttcaataa gtatagtttc taggcctttt tatctctcat 63720
tccatttgga tatttcaatt atgcaaaagt taatgtactt gattgtgtcc cataattttc 63780
ataggcctgc tttacttttt ttttttcag ctcctctgag aagttagttt caaatgttct 63840

atcttcaagc ttactgattc tttcctctgt catcaaaact atgggcacgg tggctcacgc 63900
ctataatccc agcactttgg gaggccgaga agggtggatc atctgaggtc aggagtttga 63960
gaccagtctg gccaacatgg tgaaacccca tctctactaa aaaaatacaa agaattagct 64020
gtttgtgatg atgggcacct gtaatcccag ctacttgtga ggctgaggca ggagaatccc 64080
ttgaacccgg ggggcagagt ttgcagtgag ccaagatcgt gccatcgcag tctagcctgg 64140
gcaacaagag cgaaactctg tctcaaaaaa agaaaaagga tagaggaggt tgggaagagg 64200
gaggcaggga gtcagcgtca tagcaatggg tgagaaagat gaaagatgtg tggccattgc 64260
tggctttgga ggtagaatgc gggcagcctc caaaacctgg aaaaagcaag aaagtagatt 64320
ttccctgaga acctccaaaa agaaagcagt cctgccgaca ccttgattct agcccagtaa 64380
gaagcatttt gggctttgga catccagaaa tgcaagacaa aatttgtgtt gttttaaggg 64440
attaaatctg tgatagttta ttatggcagc aacagaaaaa gaatatacct gtcaaaactg 64500
ggttcccatc ttcaaaggct tcccctctgc ttgccgtcca gtccacagct ccacttgcag 64560
gtgcagtaca gttggctttc gagcagggct tggattgaaa attcagcatg taaggaagaa 64620
gtctaaagtc cagacgaacc ttgaacaccc cttaattgcc cctcccgtcg agagcaggga 64680
ccaggatcca ctgagtgtag agcagagcaa cattcctcaa ggtgtcctca cccccaggca 64740
ccctcccaca gggctcaacc agctaacctc tttccaccct ttgaattgtc tctctctccc 64800
ttcaggcata taattaaatt tgcataaatt caaagagcat atgctgcaat cccattgtaa 64860
tttcagattt gtttttattt cctgagtttg ctaaggaaat aataaaaaca aaacctcctt 64920
aaaaaaaaaa aagatttcta atgagttttt aagttcagtt tttgtattct ttatttctgg 64980
gattaatatt tttaaacatt gtttttattt ctctttccaa tttcttattt tcttcctgaa 65040
ttattttcct tttttttag ttttctgttt tcttataatt tctttttta acatttttct 65100
ttctttaaat attttttaa attccttatt tcctcagaga tctttgttct gacctgtaat 65160
ttcttgaact tctttaagag gcttatgctg aattcattgt cagacattta atagatcttt 65220
aattgttcag ggtccattgt tgagcctctg ttgatttctt ttggtggtgt catatttccc 65280
ttggttttta taatcttcgc atatttatgt cgatgcctgt gcatttgaag agagagacac 65340
ttccagtttt cgcaggtgtt ctttgttggt cttagatctt tagttcttag tattcttttt 65400
ttttttttta agacagagtc ttgctctgtt acccatgaat gcagtggtgt gatcttgggt 65460
tactgcagcc tctggctccc aggttcttcc agcgattctc ctgcctcagc ctcccaagta 65520
gctgggatta ccggcataca gcactacgcc tggctaattt ttgtattttt agtagagatg 65580
aggttttacc atggtggcca ggctggtctc gaactcctgg cctcaggtga tccacccgcc 65640
ttggccaccc aaagtgctag gattacaggt gtgagccact gtgcctggcc agtattcact 65700
cttaaatgtt gacttgttgt tgcttctgct ggggtaggct tatagtgacc actgcaacta 65760
aatttctctt ctgtcattgt ttcctagtgt ggggaagtct tattttgtgc actgaagctt 65820
aaatactgtg ttgggactat attgctgtgc tgtcattgtt tccaattccg gaaagttatt 65880
aatcccagtc ttttaattgt tttggggcca ggctggggga ggcttcatga aagaaccttg 65940
ggtttgtgga aaatccggct aagaattctt gctgtactgc caactgtgct tcctgtattt 66000
ctatgacact aatcggtctg ttaaatgtgg catctctact aattacagtg ctgagtagcc 66060
accaggctcc atgcatcagg tactgcattc agtgtcctat tctttgttct cagttcacct 66120
cagatgattc aattctccta gcactcccaa aggttttat gagagaggac cagagtggat 66180
tactgatgaa gattcctaag ctggagggga gactgaacat ccaattccat ttccccccct 66240
cctccttagc agacataggt ctagggaaat tctttgtgag tggcattata ctagcttggg 66300
aaaggggta gtgcaaccgg aaatgaccgt ttcttttaca ctccacaact ttccttgatt 66360
ctgcacttct cctctgagtt ctgttgtatt tacaatggag ctcccatctt tgaatagctt 66420
ctagttttat ttttatgggg acagtgatgc taggggattt tctattccat catctttctt 66480
ctttctgact ttaccctcca aaagtgtaat cctcaaaatt tcaagatttt tttttgtatg 66540
ccttccatgg agctcataca acaaacaatg atacatttga aattatgaca atatgactga 66600
agcataatta gtttcctaat cccagagtcc tttaattaaa cacaagaggg atgattccag 66660
aagtaacact catcactgct attccttagt ccctctgaga tttgctttaa gagcaggtct 66720
ccacctctca catcttgcta taaatatgac tgcattttaa agcttagaga agcacctaaa 66780
aacctttta atcaaatatc aaggaaatga gtatatatca gaaccagaat taagtccttg 66840
catttttatg ttctttattt ttggttaaag gggaaatata ttaatgtatt aactgaatct 66900
cacaaaatgt ttgcattttt attattaaaa atcatttttt tcttttactt tgcatatcat 66960
tgggggggcat gaatcagttt ctcctctgaa aaattctcag tgccttctct cagaggctgt 67020
gagcaaagaa gtaggaaagg aggttagcta tgaaaagtgt taactccatt tgtaacaata 67080
ctcattacat gcttactata tgcaaatcat ttttctaagc actttaaatg agtaaactat 67140
attattcctc ctaaaaactc catttctcat tttatagctg ggaacactaa cgtaagacag 67200
tttagataat ttgacactgt ggtggtggta cgatttgtta ctgatgaagc tagaatcaaa 67260
atcactaatg gggattagat tcataaaatt gttacttttt gaatactaaa atatttactc 67320
agacattcag tatattagag taggcccagt atcatgtaat ctaatttatt ttgcaaaaat 67380
tgcaaatgag cgagatatta accaagttca attagaaagt tagttggtgg cagagctgag 67440
cataggttcc agagccaccc agaacttttg tttgcttggg agaatctctg agcacacttg 67500
gatagcttag ggaattgctg cttttgggta tcactggtca tttgcatctg atcataggta 67560
ctagactgct gacaagttgg ccattagtgg acaggtatat tgagagttga tttaagtgtt 67620

gcatctgtca aggttaacac gaaacattct gttggccctg agtctcacat ttcaatctag 67680
aatccatctg tgtactttac tttccttgca aataaatgat ctagtgaatt tatatatttt 67740
ttccacttag ctcctagcac acatcatttg tatttcaatc atttatttt cttcaccata 67800
ttagaacatt tccaatacca ttaaagaaga aggaaggcaa tgaaaaggag acggagatag 67860
cctcatacat tctcctttcg ctgaaagtaa tatcattttg aaaacagact acgggtccca 67920
tcttcttcct ctctagtttc caacagcata aagctttgta tgtctctgag gacaagatag 67980
aaaactacat tagcaaatat tttcctttgc atctccttat ttcaagtaaa tgagggactt 68040
atattttcaa atattttaag tcaccttgaa ggttgcttct taaaaaataa ctgtttatta 68100
tgcttctact cttcatttta aaaattttgc ttgtttccac ctttacaaac aaaaatggac 68160
aaaatcgtat tttatcctgc agtcctctag gacaaattta actttggggt actctgtttc 68220
ctaaaaatgc atatatgtaa tcttctattt ctagtagcag agtttgaact caattaattt 68280
ccctaacaac agactcacta aaattcgctg tgatattttt caaatagtat tttatggaat 68340
tatcactgat gaaataaaaa acaaaacaag aaagacaggg gaggagccaa gatggccgaa 68400
taggaagagc tcgggtctac agctcccagc gtgagcaaca cagaagacag gtgatttctg 68460
catttccatc tgaggtactg ggttcttctc actagggagt gccagacagt gggcacaggt 68520
cagtgggtgc gcacaccctg cgcgagcgga agcagggcga ggcattgcct cactcaggaa 68580
gcgcaagggg tcagggagtt ccctttccta gtcaaagaaa ggggtgacag atggcacctg 68640
gaaaatcagg tcactcccac cccaatactg cacttttctg acgggattaa aaaacggcgt 68700
gccagattat atcccacacc tggcttggag ggtcctacac ccacggagtc tcactgattg 68760
ctagcacagc agtctgagat caaactgcaa ggtggcagcg aggctagggg aggggcgccc 68820
gccattgccc aggcttgctt aggtaaacaa agcagccctg aagctggaac tgggtggagc 68880
ccaccacagc tcaaggaggc ctgcctgcct ctgtaggctc cacctctggg ggcagggcac 68940
agacaaacaa aaagacagca gtaacctctg cagacttaaa tgtccctgtc tgacagcttt 69000
gaagagagca atggttctcc cagcacacag ctggagatct gcctcctcaa gtgggtccct 69060
gacccctgac ccctgagcag cctaactggg aggcaacccc cagcgggggc agactgacac 69120
ctcacacggc cgggtactcc aacagacctg cagctgaggg tcctgtctgt tagaaggaaa 69180
actaacaaac agaaaggaca tccacaccga aaacccatct gtacatcacc atcatcaaag 69240
accaaaagta gataaaacca caaagatagg gaaaaaacag agcagaaaaa ctggaaaccc 69300
taaaaagcag agcgcctctc ctcctccaaa ggaacgcaat tcctcaccag caatggaaca 69360
aagctggacg gagaatgact ttgacgagct gagagaaggc ttcagacgat caaactactc 69420
caagctacgg gaggacattc aaaccaaagg taaagaactt gaaaactttg aaaaaaattt 69480
agaagaatgt ataactagaa taaccaatag agagaagtgc ttaaaggagc tgatggagct 69540
gaaaaccaag ctcgagaact acatgaagaa tgcagaagcc tcaggagccg atgcgatcaa 69600
ctggaagaaa gggtatcagc aatggaagat gaattgaatg aaatgaagtg agaagggaag 69660
tttagagaaa aaagaataaa aagaaatgag caaagcctcc aagaaatatg ggactatgtg 69720
aaaagaccaa atctacgtct gattggtgta cctgaaagtg atggggagaa tggaaccaag 69780
ttggaaaaca ctctgcagga tattatccaa gagaacttcc ccaatctagc aaggcaggcc 69840
aacatccaga ttcaggaaat acagagaacg ccacaaagac actcctccag aagagcaact 69900
ccaagacaca taattgtcag attcaccaaa gttgaaatga aggaaaaaat gttaagggca 69960
gccagagaga aaggtcgggt taccctcaaa gggaagccca tcagactaac agcggatctc 70020
tcggcagaaa ctctacaagc cagaagatag tgggggccaa cattcaacat tcttaaagaa 70080
aagaattttc aacccagaat ttcatatcca gccaaactaa gcttcataag tgaaggagaa 70140
ataaaatact ttacagacaa gcaaatgctg agagattttg tcaccaccag gcctgcccta 70200
aaagagctcc tgaaggaagc gctaaacatg gaaaggaaca accggtacca gccactgaaa 70260
aatcatacca aaatgtaaag accatcgaga ctaggaagaa actgcatcaa ctgacgagca 70320
aaataaccag ctaacatcat aatgacagga tcaaattcac acataacaat attaacttta 70380
aaagtaaatg gactaaacac tccaattaaa agacacagac tggcaaattg gataccaaga 70440
gtcaagaccc atcagtgtgc tgtattcagg agacccatct cacgtgcaga gacacacata 70500
ggctcaaaat aaaaggatgg aggaagatct accaagcaaa tggaaaacaa aaaaaggcag 70560
gggttgcaat cctagtctct gataaaacag actttaaaac aacaaagatc aaaagagaca 70620
aagaaggcca ttacataatg gtaaagggat caattcaaca agaagagcta actatcctaa 70680
atatatatgc acccaataca ggagcaccca gattcataaa gcaagtcctg agtgacctac 70740
aaagagactt agactcccac acattaataa tgggagactt taacaaacac cccactgtca 70800
acattagaca gatcaacgag acagaaagtc aacaaggata cccaggaatt gaactcagct 70860
ctgcaccaag cgtacctaat agacatctac agaactctcc accccaaatc aacagaatat 70920
atatttttt cagcaccaca ccacacctac tccaaaactg accacatact tggaagtaaa 70980
cctctactca gcaaatgtaa aagaacagaa attataacaa actatctctc agaccacagt 71040
gcaatcaaac tagaactcag gattaagaat ctaactcaaa accactcaac tatgtagaaa 71100
ctgaacaacc tgctcctgaa tgactgctgg gtacataacg aaatgaaggc agaaataaag 71160
atgttctttg aaaccaacga gaacaaagac acaacatacc agaatctctg ggacacattc 71220
aaagcagtgt gtagagggaa atttatagca ctaaatgcct acaagagaaa gcaggaaaga 71280
tccaaaattg acaccctaac atcacaatta aaagaactag aaaagcaaga gcaaacacat 71340
tcaaaagcta gcagaaggca agaaataact aaaatcagag cagaactgaa ggaaatagag 71400

acacaaaaaa cccttcaaaa aattaatgaa tccaggagct gtttttttga aaagatcaac 71460
aaaattgata aaccagtagc aagactaata aagaaaataa gagagaagaa tcaaataggc 71520
gtgataaaaa atgataaagg ggatatcacc accgatccca cagaaataca aactaccatc 71580
agagaatact acaaacacct ctatgcaaat agactagaaa atctagaaga aatggataaa 71640
ttcctggaca catacactct cccaaactaa accaggaaga agttgaatct ctgactagac 71700
caataacagg atctgaaatt gtggcaataa tcaatagctt accaaccaaa aagagtccag 71760
gaccagatgg attcacagcc gaattctacc agaggtacaa ggaggagctg gtaccattcc 71820
ttctgaaact attccaacca atagaaaaag agggaatcct ccctaactca ttttatgagg 71880
tcagcatcat cctgatacca aagccgggca gagacacaac caaaaaagga attttagacc 71940
aatatccctg atgaacattg atgcaaaaat cctcagtaaa atactggcaa accgaatcca 72000
gcagcacatc aaaaagcttg tccaccatga tcaagtgggc ttcatccctg ggatgcaagg 72060
ctggttcaat atatgcaaat caataaatgt aatccagcat ataaacagaa ccaaagacaa 72120
aaaccacatg attatctcaa tagatgcaga aaaggccttt gacaaaattc aacaacactt 72180
catgctaaaa actctcaata aattaggtat tgatgggacg tatttcaaaa taataagagc 72240
tatctatgac aaacccacag ccaatatcat actgaatggg caaaaactgg aagcattccc 72300
tttgaaatct ggcacaagac agggatgtcc tctctcacca ctcctattca acatagtgtt 72360
ggaagttctg gccagggcaa ttaggcagga gaaggaaata aatggtattc aattaggaaa 72420
agaggaagtc aaattgtccc tgtttgcaga cgacatgatt gtatatctag aaaaccccat 72480
catctcagcc ccaaatctcc ttaagctgat aagcaacttc agcaaagtct caggatacaa 72540
aatcaatgta caaaatcaca agtattctta tacaccaaca acagacaaac agagagccaa 72600
atcatgagtg aactcccatt cacaattgct tcaaagagaa taaaatacct aggaatccaa 72660
cttacaaggg acgtgaagga cctcttcaag gagaactaca aaccactgct caatgaaata 72720
aaagaggata caaaccaatg gaagaacatt ccatgctcat gggtaggaag aatcaatatc 72780
gtgaaaatgg ccatactgcc caaggtaatt tacagattca atgccatccc catcaagcta 72840
ccaatgactt tcttcacaga attgcaaaaa actactttaa agttcatata gaaccaaaaa 72900
agagcccgca tcgccaagtc aatcctaagc caaaagaaca aagctggagg catcatggta 72960
cctgacttca atctatacta caaggctaca gtaaccaaaa cagcatggta ctggtagcaa 73020
aacagagata tagatgaatg gaacagaaca gagccctcag aaatatcgcc gcatatctac 73080
aactatctga tctttgacaa acctgaggaa aacaagcaat ggggaaagga ttccctattt 73140
aataaatggt gcagggaaaa ctggctagcc atatgtagaa agctgaaact ggatcccttc 73200
cttacacctt atacaaaaat caattcaaga tggattaaag acttaaacgt tagacctaaa 73260
accataaaaa ctctagaaga aaacctaggc tttacctttc aggacatagg catgggcaag 73320
gactttatgt ctaaaacacc aaaagcaatg gcaacaaaag ccaaaattga caaatgggat 73380
ctagttaaac taaagagctt ctgcacagca aaagaaacta ccatcagagt gaacaggcaa 73440
cctacaaaat gggagaaaat tttcacaacc tactcatctg acaaagggct aatatccaga 73500
atctacaatg aacttaaaca aatgtacaag aaaatcaaac aaccccatca aaaagtgggc 73560
gaaggacatg aacagacact tctcaaaaga agacatttat gcagccaaaa aacacatgaa 73620
aaaatgctca ccatcactgg ccatcagaga aatgcaaatc aaaaccacaa tgagatacca 73680
tctcacacca gttagaatgg caatcattaa aaagtcagga aaaaacaggt gctggagagg 73740
atgtggagaa ataggaacac ttttacactg ttggtgggac tgtaaactag ttcaaccatt 73800
gtggaagtca gtgtggtgat tcctcaggga tctagaacta gaaataccat ttgacccagc 73860
catcccatta ctgggtatat acccaaagga ctataaatca tgctgctata aagacacatg 73920
cacaagtatg tttattgtgg cattattcac aatagcaaag acttggaacc aacccaaatg 73980
tccaacaatg ttagactgga ttcagaaaat gtggcacata tacaccatgg aatactatgc 74040
agccataaaa atgatgagtt catgtccttt gtagggacat ggatgaaatt ggaaatcatc 74100
attctcagta aactatcgca agaacaaaaa accaaacact gcataccctc actcataggt 74160
gggaattgaa caatgagaac actggacaca ggaaggggaa catcacactc tggggagtgt 74220
tgtggggtgg gggaggggg agggatagca ttgggagata aacctaatgc tagatgacga 74280
gttagtgggt gcagcacacc agcatggcac atgtatacat atgttacaaa cctgcatgtt 74340
gtgcacatgt accctaaaac ttaaagtata ataataataa agaaatgggt cttctctacc 74400
caagaacaca cacacacaca cacacaaaaa aaaaaaaaaa aaaacaagaa agacaaacag 74460
tatgctcatc acacacaaat gcaaggtcat ttttaggcat tagatggggc ttaatttaat 74520
ctaagttttt catgtacttt gtatcttcta acctgcatac ctctatttcc tcttctagcc 74580
ctctacctct ggtaaccact gttttatttt tatcactgta tatttaattt tttaaatatt 74640
ccccatataa gtgagatcat tcagtatttg tctttctgtg tctggtttat ttcacttaga 74700
ataatgtcct ccaggcttgt acgtgttgta tcaaataaca cgatcttcct tttcagggat 74760
gaataatatt ctattgtaaa tatataccac catttcttta tctatttgtc cttcatgagg 74820
cacttgggtt gtttcaatac cttagctgtg atgaataata ctgcaaaaac atggaagtac 74880
agatgcttta agaggtggtg aattcacagg ataaacaagt ctagagatct aatgaacaac 74940
atgaggacta ggggtaataa aattataccg tatttgggat tcgtgctaaa tgaatatatc 75000
ttagctgatc ttgcacacag acacacaaaa agggtaacta tatgttaatg tttgtgttaa 75060
tttgctttac tatagtaaga ttttactcta tgtatcacat gacattatgt tgtgaacctt 75120
aaatagatgc attaaaattt atttttttaaa agggttcatt aactgcggtt caggcaatct 75180

ataaaccttt tacattttgg aggtgagatt ttcagtatct gcatttatct tggattttca 75240
aggggttcat ttgaacaagc aaacaaacga agaaacaagc ttaagaatct ttcttacaat 75300
gaattcttcc taagtagttt tcttgactgg tttcttttct ttttctgaaa cagtcttttc 75360
ttattcactt gtgattgaat agcctacaaa attgatctag acaattaggc tcttgaaatt 75420
gattacggcc tagtgcattc tcacctggaa taaaagcctt cctttggagc atcttctatt 75480
ttttaattgt tgagagttgg ggactttcct tctcattctg tattttcagt gtcagttttt 75540
atggcttgtg tgttccaggt atggttgggt catggtcatg cagtctattc acaaggcctt 75600
gtgcttagaa gaccatccca agcttggttt aatgctttga tctcatggtc ctgaaatttt 75660
acataatttt tttgtacaag gattcttgca gttttatttc gtattgaaat aggagagttt 75720
ccttatcctc ctcgcatggc aagtgacagg agtgtggctt gcttctttgg tgccccactg 75780
ctcaaatccc taggggtgga atgcagacag tcaggtgtac aggtcgtggg agtgtttttg 75840
ggctccaacc ccacagcagc gtctagggtt aagtgtttac agctcctgaa gccccaatgg 75900
gtgtgtgtta cagtgtgatc tttcagcttt gctgtcttca ggcggcttgc attaatcagc 75960
tcagttagac cctctgcatt accacaaggg cagcaggctt tctgtattcc aagttcttgc 76020
ccagtgtatc ggaaaaattg gatcacatgt ggacatggag gatgagtgca aggttttatt 76080
gagtggtgga ggtggatctc aatgagatgg atggggagcc agaaagggga cggagtggga 76140
aattggtctt cccctggagt cgggcagctc atttgccgat gatgctctga ccacccccag 76200
acgaactccc ctcggtgtct gcatcgttcc accatcactg gtctgctggc atctgctggt 76260
gtctgctgtt gtgctcctct gctcctgtca acatccagcc acttgtgtcc atgcccacta 76320
aggcctcggg ttttttgggg cacaggatgg ggcctgtggt gggccagagt ggtcttggaa 76380
aatgcaacat ttgggtgcaa aaacaggaga gcctgttctc acttaggtcc ttgggcaaag 76440
gtctgtgggt ggagccctca ccagggaccc cacccttctc tacacatcac ttcccaaccc 76500
tgctcctgta tcaatattgg attctgtata ctatgtatct gaggtcccag ttctaggttt 76560
tacaacttgt aacttttttcc tatccttgac ctactcctct gctcgttctt cctctatcat 76620
gcagaacctt cccaacatac tgacctagca agcccaggtg ttccttattc aagaggtgtt 76680
caccttttat tctgtcctga cacaggggtt agacctactg tggttttcct tctttgtaca 76740
cacatcttag ctgtagtctt tttacctctc tccacttatg agcccttctt ctctactgca 76800
attgtgtatt tctggatccc ataaaagcca atgagcctct cagtcaaaac tgtttcagat 76860
ctcacaggca gttattaagc cacatgcaac aaataagtta ttctgcaaat gaactcctct 76920
agtccaattt cctactgtgt ctttttmaat ctgaaattcc cacatcctga tatattttct 76980
ctcatttttc tttttgtaac cactgaaact gtactctggt gaaggccact gatactcttg 77040
ataggaactt gaaatgagta acacttaacg tgagaaagat ttgaagtgaa taagagtaaa 77100
gtgaaatatc tgtagtttag ggagcttagc tggattttct gcactttgca ttaagtgtgt 77160
gtgtgtgtgt gtttgtgtgt gcacatacac acacaaattt cactcagtga tagtttccaa 77220
agaaccatta catctggtat ccttgtattt gctcgttatc cttctgtaga cccaggaaat 77280
ctgagacagg tctcagttag tttagaaagt ttattttgcc aaggttgagg acacacctgt 77340
gacacagctt ccagaggtcc tgaggacatg tgttgaaggt ggtcatggta cagcttgctt 77400
ttatacattt tagagagaca taatctatca atcaatacat gtaagattta ccttggtttg 77460
gtgcagaaca actcaaagtc caggttgggg cagggtggcg tgtcgtttcc aggctgtagg 77520
taaatttaaa cattttctgg ttgacaattg gtttgtctaa agacctgaga tcaatagaaa 77580
agtctgtgtt atgttaatgc tggagaggaa taatgaggca tgtccgaccc ccatttctct 77640
taaccgcctg aactagtctt tcaagttaaa ttttacaagc cctggctgag gaggaagtcc 77700
ctttatttat ttttggttta cacttcttat gatatccttc ttagaagttt tcttgctaca 77760
gttgttgcca taaactttac ttttttttttt tttttggtag aaccttattc tgtcgccagg 77820
ctggagtaca gtggctgatc tccgctcact gcaacctctc catcctgggc ttgagcaatt 77880
cttctgcctc agcctccgaa gtagctgaga ttacaggcac gtgccaccac acctgactga 77940
tttttgtatt tttagtagag atggggtttc tccgtgttgg ccaggctggt cttgaactcc 78000
tgacctcagg tgacccacct gcctcagcct cccaaagtgc tcgaactaca ggcatgagcc 78060
atcctgcctg gccaaacatt actttttttaa tgaggtcttc cttgactttc ctttccccaa 78120
aactgtattt cctttatttt tctccacaac agttatttga atattattta tttacttgaa 78180
ataatagctc tctctgtcca ctagaatggg aaaaccaaaa aaagcagaaa tttttgaaag 78240
gtttgttcgc ttctgcatct atagtgccca atcacacaaa atattgagta tcagtaaata 78300
cttgtttcat gaatatctat tttgatgccc aaaatgacat gagaagagga aagataaata 78360
ttattatttc tatttacaca tgggaaaaca cagtgattca ttcagttgca aagcttttaa 78420
attactgaag taactctgta atccagggct cctaaccagc tcctaacatc aagatatgtc 78480
atagcttcta cttgatattt cttagatata tcacatacag gtactcaagg gcacaagtca 78540
gtcacatagt agaacatgac taagtgtaag gagaatgtgt ctggcacaat cctgtttatc 78600
atggtgcact atgtaagaat gccatgacta tctgtgtccc cagaagttcc ttatcatcat 78660
atatctcaag gccttccagg tctatacata caattttatg aaacagttta gctttagaaa 78720
ggtgatatgt attgcagcag cagctaataa aggaagtttt taaaatacat aattttcagt 78780
tcttacctca gattctcaga ttgtatgcac tagtagtgtg acctcggttc aattatttaa 78840
ccttttgta cgttgctttt ctcaatagta acagcaagac agagttttga ttataggcat 78900
tctaaaactt acagtaaaat aaatgaggtg actcattttg agatgagaca agaattgaga 78960

agcctgtaat aaaaatttca agtatgaaag aatgggcaaa aattaaaaac ttctcgaaag 79020
aaaggaaaat gtgaattagc aaagattctg atgtgaagag aggtaaaaag ccaaaaggat 79080
ttgtaatacc tcttgttaga tacaagttgt gattaaaaga ttaaataaca actctgaaaa 79140
tattaattta ttttctttct aaacacatta agtcattgct ctaaacatcc tcaagatact 79200
atatgaaaat ttattttatt tcacagttct ttgttcattt atttggcata tgtttattga 79260
gccattctga ctgtgcattt tgagctatga tcttggaaca cttgcattag aatcctacag 79320
ataatcatta gagatgaaga ttccagtacc ccacataaga atttctgcaa aaacctgtga 79380
tatgggctaa taaacaagct tcccaagaat tgcttctaca ctaagtttca ataatcactt 79440
aattctcctg agaacatgac atcacgcaag gctacatgga gataaatctg tgagtaaaaa 79500
ttgcaagtca ttcattctta catgctctcc tttttgcatt ttaagaaaat ttagtataaa 79560
aatatgtgtt gagaagtata attaaaaaat gaacttcact ttgtccctgc tctcaacacc 79620
attatgtgat gagaaagtat atactttata atgatacaat gggggctaaa taataaatta 79680
gagaaatata gaatactctg gaaataaaga agagagaatt cagtgacacg tattaggaga 79740
aatggatgag aaatgcttca caaagaagtg gcctttaaag tgggtcatgt agcatgaata 79800
gaattaggtt ttaaaaagaa caggaagata gaagagatgt agctcagtca tatagataga 79860
tagatagata gatagataga tagatagata gatagataga tagatgatgg atagatagat 79920
acaggtgtag gtatagatat atctataggt gtagagaaag aatatattag taatgtaatg 79980
cctaggcaaa tgatataatg acattgtaga cttacaaaat tatagtaagt ctaccataat 80040
taccaaaacc aaagtttatg gtagtaccca gaaatagata actttcgatg atatatgaag 80100
atgtggagaa aaactaattc aaatgaacat taacatggta tgtgcatttg tttatattgg 80160
ctttatttcc atagcttgaa gagcaaactg tcaggaaatg tccaacacaa atggcagtgc 80220
aatcacagaa ttcattttac ttgggctcac agattgcccg gaactccagt ctctgctttt 80280
tgtgctgttt ctggttgttt acctcgtcac cctgctaggc aacctgggca tgataatgtt 80340
aatgagactg gactctcgcc ttcacacgcc catgtacttc ttcctcacta acttagcctt 80400
tgtggatttg tgctatacat caaatgcaac cccgcagatg tcgactaata tcgtatctga 80460
gaagaccatt tcctttgctg gttgctttac acagtgctac attttcattg cccttctact 80520
cactgagttt tacatgctgg cagcaatggc ctatgaccgc tatgtggcca tatatgaccc 80580
tctgcgctac agtgtgaaaa cgtccaggag agtttgcatc tgcttggcca catttcccta 80640
tgtctatggc ttctcagatg gactcttcca ggccatcctg accttccgcc tgaccttctg 80700
tagatccaat gtcatcaacc acttctactg tgctgacccg ccgctcatta agctttcttg 80760
ttctgatact tatgtcaaag agcatgccat gttcatatct gctggcttca acctctccag 80820
ctccctcacc atcgtcttgg tgtcctatgc cttcattctt gctgccatcc tccggatcaa 80880
atcagcagag ggaaggcaca aggcattctc cacctgtggt tcccatatga tggctgtcac 80940
cctgttttat gggactctct tttgcatgta tataagacca ccaacagata agactgttga 81000
ggaatctaaa ataatagctg tcttttacac ctttgtgagt ccggtactta atccattgat 81060
ctacagtctg aggaataaag atgtgaagca ggccttgaag aatgtcctga gatgaaatat 81120
tgtcatgacc atggtgatgc ctttgtttcc taataaacat taaatcgaaa tctttggctc 81180
acatgtccta gcgttctgat ggtgagtttt aatattctct gtgagtctat gttgagtgtc 81240
tcagctaaaa agctcatgct gggtaaaaat gagatttttc taggctttgc tcctccacat 81300
atattccatg aatcagcagc atgagctctt ccttggaggt tgttacacgt acagaatcaa 81360
agtctgcacc tcaggtgcac tgtatttaaa tatgtgtttt atccaaactc ctagatgatt 81420
gataagcaca ctgaattttg aggagcactg ctgtgggtga aacgtggcat gccctggaac 81480
actgttgtgc tctttttgtt tacaacggca aacaaaataa atgtgctccc agcccaattt 81540
cttgaatgta ttctattctt attctcgcct gctgcttcag cagagatgtc tttaagaaac 81600
ccattcttct gcactccaag aaatccattc ttctgtactc ctttcctgac ttgctgtggt 81660
agaccagaaa ctaggggcca tggtgatggc ttctggtttt tataagtgct cttacatagt 81720
gaagagtggc aatggaaaaa gaggggaaga gaatgattgt acttttctta aacttgagtt 81780
tatggatcca gctctctgag ttacataaac ccaccttccc attctgagcc tccagtgtat 81840
ggaggtcact ttctgtttcg cctgattttg aaaaatttaa gcaggctgat ctgagaataa 81900
ggactaaaat taaaatggga aatgaaagaa ccgagtaaac aataggtcat acgatagagt 81960
atccaccttg ctttgaatat tccttctatt tgtaattaag tttaggtaag taatgttaat 82020
tttttgtttt ccttagcact ggcctgttta tatgtgtcca aggaagtaag gatttcttcc 82080
acaggaagag aagacattgc tctggcatct tcaagaaact ttagaaagta gaaagaattc 82140
ctctttagac attatgctaa aaagaattat ttaaggaaca agtaaaaaat atatttttt 82200
aagatagagt ttcactctgt tgcccaggtt ggagaacagt ggcatgatca tattaataac 82260
ttcctccaac tcctaggctc aagtgaccct cccgcctgag cctccctagt agctgggacc 82320
acaggtgtgt gctaccacac tcggctaatt tttaaatttt tttgtagatc tgatatctgt 82380
ctatattgcc caggcaggtc ttaaactcct gggctcaagt ggctctcctg cctcagcctc 82440
ccaaagtgct gggattatag gcataaatta ctgtgcctgg tcaagattct tctactctcc 82500
aaactttcta cttattgcct caaaatacta aatattttct aaatatcttt attggaaggc 82560
aaattctgcc attacgccat tagaaaataa aaagataaaa tattttgtgt atatattaag 82620
ttgtacatta agtaactgta tttataaggc ttctgcattt tgttcctgtt gttgtttggg 82680
tttgttgtta gttttacaag gaaaaaagag ccccaagcta tctaagtcaa caacggcaat 82740

ttatgacaaa tatatggtta tatttcacag cttctaaaga caagagtaca gctgaatctc 82800
ggtaatcaac caaaatgctg aactctaagg ccaagagaaa ccgtatgtct cattgtatat 82860
ttcatatcca tttgttacta ggatacataa taattgagct taattatgac ccttgcatcc 82920
agcaaccctg ctcatttgtc ttaattagag tagaattttt gtttgtttct tctataaatt 82980
ctctaaaatt tacttcattt catctgcaaa ttaacagatt ttattccttt ccaatattta 83040
tgagttttat ttatttttcc ttcatcattg cagacttaga ctgatattat aatgtcaaat 83100
agaggtagtg aaaacaaata tcaaggcctt gttcctcatt tcatttgtgt tttgtttctt 83160
tgtttgtttt gagacagagc cttgctcttg tcgcccaggt tggagtgcaa tggcacaatc 83220
ttggctcact gcaacctctg cctcccgggt tcaagtgatt ctccttcttc agcctcctgg 83280
gtagctggga ttacaggcac ccaccaccat gcctgactaa tttttgtatt tttaatagag 83340
atgggttttc gccatgttgg tcaggctggt ctcgaactcc tgacctcatg atccacccac 83400
ctcagcctcc caaagttctg ggattatagg cgtgagccac catacccagc cttgctcctt 83460
attttagggg gaacatattt aatgtatcat caggttgtat atcattagtt tgatttttt 83520
tgtggcatta ttaagattcc atttaaaaat ttgacatatt tgttcacgca tctgttgtta 83580
gaagtgggat tgttttatta agtatcaaat tatcttaagt ttcaggctta atgacagtgt 83640
aatttcaggg gcataaggaa gattttccca ctcatccaac atagacttat actgagagag 83700
atgtaattac aattaacaac catgtatttc taatttcaca tgtgcgggat gctttgttct 83760
ctttttctca tctgatcaag gcaccatggc tccaacatcc tcatctatgg gatatcttgc 83820
tgacttttgt aggtagaagt aagcattttc tgcattgtgt ccataatcta ctttttaatt 83880
agaaatttta cttaatttca cactatattc atttctgtgg tcaaaaaaat caaataaaat 83940
tgtgaattgg ggtggtggct gaaggatagt aataatatgt ttaggcaatt tgggtaaatc 84000
atggagtatt ttttctttaa tgaagaaata tgctgttctg cttcatccat ttattgcaat 84060
tccaaaaagc aaactaacta tagtcagatt atttggattt tatgagaatt tcataatatt 84120
ataattattg caaaatttat taattttaaa aaatcacgtg aaccaaaact aaacacttgt 84180
ggatataatc aagattcttg ttttcaatat atttactcaa aaaacattta tggataactt 84240
agcatgttcc aggcagtgta ctgagtgttg tgtatacagt gaatataaaa tacagtttct 84300
tcttctcttg gataacaggc tggaggggaa gatatatctt aaacaaacaa aataaactgc 84360
acttaaaatt aagattgtga caagtgtcat aaataagtag aataaggcaa caggagagag 84420
aacaattatg atggtgcaca gattttactg taggtagaag atgtcttctc ttagggttga 84480
tatttaagct ggcatcctag agatgattag ggaataacta gaagaagttt gggagaaata 84540
gaattcaaag cagtaattag aactaagcac cacatcgaaa tgatagaaag atctcaacaa 84600
cttaacataa caactaaata atataaggag gctgctagct agactaatat gaagaaaaga 84660
gagatggccc aaaaaacaaa attataaatg accaaggaga tgttaccact gaacccctag 84720
aaatacaaat aaccatcaga gactattatg agcacctcta tgcacacaaa cgacaaaatc 84780
tagaagggat ggataaattc ctggacacat acccttcaca agactaagcc aggaagaaat 84840
tgatttcctg aacagaccaa taatgagctc caacattgaa tcagtagtaa gtagcctacc 84900
aaccaaaaga aatcccagta ccagatggat tctctgccaa attctaccag atgtacaaag 84960
aagagatagt accatttgta ttgaacctat gcaaaaaatt taggaagaat gattcttccc 85020
cagtcgttcc ttaaggccag catcattctg atatcaaaac ctggcaaaaa cacaacaaaa 85080
aatgaaaatt ttaggccaat atccttgatg aatattaatg caaaaatctt caacaaaata 85140
gttgcaaact taatccatga gaacatcaaa agcctaatct atcacaatca agtaggcttt 85200
atccctggga tgtgaggttg gttcaacata cataaagtta aatgtgattc atcacacaga 85260
cagaactaga ggcaaaaact atgtgttatt tcaacagata cagaaaaggc ttttattaaa 85320
attcaacatc ccttcatgtt aaaaactctc agtgagctag gtattgaagg aacatacctc 85380
aaaataataa gagccatctg tgacaaaccc acagccaaca tcacacagaa tggggaaaag 85440
ttggaagcat tcccccttaaa aacctgcaca aggcaaggat gctgtctctc accactccta 85500
tttaacatgg tattggaagt gttagccata acaattaggc aagagaaaga aataaagggc 85560
atccaaatag gaagagagga attcaagcta tgcctgtttg cagtcaatat aattttatat 85620
ctagaaaacc caaaaacttc ttcagctgat aaacaacttc agcaaagtct cagggtacag 85680
aatcaatgta caaaaatcac ttgctttcct atacaccaat aactgtcagg caaacagtca 85740
aatcaggaat gcaatccaat tcacaattgt tagaagaaaa ataaaatact taggaataca 85800
actaactaga aagatgaaag agcactacaa tgagaattac aaaatctgct caaaaaaatc 85860
agagatgaca caaacaaaag ggaaaacatt ctgtgatggt taataatgag tgagtgtcaa 85920
cttgattgga ttgaaggatg caaagtattg atccttgatg tgtctgtgag ggtgttgcct 85980
aaggagatta acatttgagt cagtgaggtg ggaaaggcag acataccctt aatctggatg 86040
ggcacaatct aatcagctga cagtgtgccc agaatataaa gtaggcagaa aaacatgaaa 86100
aggttagact ggcttagcct cccaatctac atctttctcc cgtgctggat gcttcctgcc 86160
ctcaaacatt gaactccatg ttcttcagct ttgagacttg gactggcttc cttgcccctc 86220
agcttgcaga tggcctattg taggaccttg ccttgtgatc atgtgagtta ataatactta 86280
acaaactccc atatatatat atatacacac acacacacat atatatacat atatatacat 86340
atatatatat gtatatccta ttagttctgt ccctctagag aaccctaata cacattacat 86400
gctcatagat aaggaaaatt aatattatta caatggccat actacccaaa gcaatttaca 86460
gattcaatgc tattcctgtc aaactaccaa tgacattctt catagaatca gaaaaaaagc 86520

tatttaaaaa tttataggga accttgaacc caaatagcca aagcaatcct aagcaaaaag 86580
aacaaacctg gagcatcacg ttacctgact tcaaactata ctataaggct acaggaagca 86640
aaacagcatg gtactggtgc aaaaacagac aaataaaaca atgaaacaga attgaaaggc 86700
cataaataag accacacacc tacaaccatc tgatttttga caaagctgac aaaagaacag 86760
ggggaaagga caccctattc aataaatggt gctaggctga ctggctagcc atatacagaa 86820
gattgaaacc ggacctgttt cttacaccat acacaaaaat caactaaaga tgaattaaag 86880
acttaaatat aaaaccctga actgtcaaaa ccctgaaaga caacataggc aacaccattc 86940
tggacatagg aactggcaga gatttcataa tgaagacacc aaaaacaatt gcaataaaag 87000
caaaaattga caaatttttg gatataaata aacttaaaag cttctgtgaa agaaactttc 87060
aacagagtaa acagacaacc tacagaatgg aataaaatat ttgcaaacta tccatctgac 87120
aaaggtaata tccagcatct ataaggaact taaacgaatt tacaagaaaa aaatgatccc 87180
attaaaaagt gagaaaagga catgaataga tacttttcaa aagaagacgt aaatgggagc 87240
taaataatga aaacacatgg acacaaaggg gagaacaaca gacactgggg cctacttgaa 87300
gatggaagat gggagtaggg aaaggatcag aaaaaataac tgtttggtac taggcttagt 87360
atgtgggtga tgaaataatc tgtacaacaa acccccatga cgtgagttta actgtatcaa 87420
aaaccttcac atataccccc gaacctaaaa aaaaaaaaaa aaaaaaaagg ttaaaaaacg 87480
gctctgtgaa ggttctgagt ttgaaaaggg cttgacccaa gaccagtatg gctataccag 87540
ggaatgaggg agacagtgac agataatgct gatttaggtg tagactatag taaaagtttt 87600
ttttttttt ttttttttt ttttttgagg cggagtttcg ctctgtcgcc caggctggag 87660
tgcagtggcg cgatctcgac tcactgcaag ctccgcctcc cgggtttacg ccattctcct 87720
gcctcagcct cccgtgtagc tgggactaca ggcgcgcgcc accatgcccg gctaattttt 87780
gtatttttag tagagacggg gtttcaccgt gttagccagg atggtctcga tctcctgacc 87840
tcgtgatccg cccgcctcgg cctcccaaag tgctgggatt acaggcgtga gccaccgcgc 87900
ctggcctata gtaaaagttt ttatttcggt atatcattgg aaaaatgtgc ttccaatgaa 87960
gcattcatgt atctcttttg ccttcataat tagggcaaat acatgtacta tttgatcaca 88020
tgtacaggtt acatcttata ttcatatgtc ctagctaagt gaggattgga tacttggtaa 88080
aagggaaaaa aaaggtaggc aaaatatttt gatctttatt ctgtttgttt tcataggttg 88140
gcaaataaat aacattttgc taagttacac attcaccagg ttttccataa aatattactt 88200
cccaacacct ggattaaata ccgtgaacta actcaactac aacaattttg aggtccaaat 88260
tattttaccc agtagatttt acaatttaac ttctttattc taaccatttt gtgaaaggta 88320
gtgtccaaca cacaattcag gctgaaagag aatcacagaa acatcaacgc ttatgaatta 88380
tgtcattatt tttttacctt tattaatcat tccattcatc tgattggaat aatgaaccat 88440
aatgcaaaac cccatgtgaa taaactcctg atccccaaag caatcatctc gtagtgatta 88500
caaattgttg caatcaagtt gcttaagtcc caagtttcta ttgagcacat gagtctgtac 88560
caattactac attttccctg gagttgtgtt ttttagtaac attgcaattg tttcttcaga 88620
gtgtattaaa aaatacttta gtgataagag ctccaaatta caagatgggt ataacttatt 88680
cttgttaact catttttgg agcactagtg tctgtgggag tacctttctt ggcatcaggt 88740
aggaataaaa attctgactt ctgcacttgt actttctctt cttgtctcaa ttttggattt 88800
ttatcagtag gagtcaagtg acaataaaag tagcttcaaa tgttagttta gattctcagc 88860
ctattaaaat gagtttgtat acatttctta tctagacggg tgatgagaaa gtattacact 88920
ttaattgaaa ttattcacaa caactgtttc actttacaaa tatcattctc accattcatt 88980
caaatattga acccacttga ataactattg gagtaaacat taatatgaca ataaattcaa 89040
gtgttgaaca caccttgaat aaccactgga gtaaatatta ataagacaat caattcaagt 89100
gtttactaga ataatatctg ttgaaaagac cccttataaa ggaaacgctt tatcaaaagt 89160
tcaaaagaga tcttatttt ttatggctga atagtatttc attgtatcat agatggacac 89220
aatgaaatat catagacatt ttacattaag taaaataagc caggaaaaga atattaaata 89280
ctctatgttc tcactcgtac gtggaaacta aaaaatgttt atgtcagaga agtaaaaagt 89340
agaaaagagg atactagagg cttcctagag gctaggaagg gtagagagaa ggaaggatag 89400
gcagaaattt gttaaaggat acaaattata gcaaggtaag agggataagt tctactgttc 89460
tatggtcctg taggattact gtagttcata atatatagtt tcaaatagct agaaaaagga 89520
tagtgaatgt tcccataata aagaaatgat aaatatttga cataagggat atgctaatta 89580
cccttatctg atcactatac atcatatgca tcaaaacatc actgtgtacc cactaaagat 89640
gaacaagtat tatatgtcat ttaaaaataa agtaaaataa aggggaaaaa tttgaaagag 89700
ggacagatca attcattaga aattgtactt tttctagaaa acataaaaca cttatagaac 89760
tttacatatc tttctttaaa ttatcagaga aaatttggag caaatttaat attaactaag 89820
aagtacatga ctagtgatga gtgtagaata gaaaaatcta tgaatgcaaa atcagaatgt 89880
tagaaatgat gggcacttta gaaactattt tgtcaaagtc tgacattata tagattagaa 89940
ttctaaagct cataggaaaa aatcaccata tcgaggtaaa gagactgctt tctaaatgaa 90000
gacttaaatg agtttttatt gacatatcat aggcaagaat aaagccacaa catataacat 90060
gccctagaac ttaggttcat gcactgtgct tagataatgt gccctgttat tttatttgtg 90120
agacactaga atggaaaaca aagagaaagg gatggtgctg aagtccattg tgactgtgta 90180
gacaaggaag gtgctcctgg ctgagaaact gatattcaca ggagttgctt ggctccacca 90240
actactagac ttatctacct ctgttttaat tgaagatcgt aaaatatcat tatttgccca 90300

gcagttcttg cagacttatc tatgtcttat tgctcagttg aggtcgttcc tgggctcaat 90360
aactgaaaat aagcttcagt aatctttacc tctgttttct atggtctttt ctttggcttc 90420
aatttctcct agtctcccac acaattcatt ttataggatg acataatggt gacctgatgt 90480
atacagaatg tttagtaact atgaaagcag ttccgtgcta tcttattttg ctatctcaaa 90540
tatttgtatt ctattttatt tcattttatt ttatgagaca ggatctcact gtgtcaccca 90600
ggctaaagtg cagtggtgca atcatggcat actgggccca agggagcctc tcacctcaac 90660
ctccccagta gctgagacca caggcatgca ccaccacacc tggagatggt ttttttttt 90720
aatatttgtt tttgtagaga tggagtctcc ctatgttgcc cagttttgtc ttgaactctt 90780
aggctcaagc ctcagcctac caaagtgcta ggagtaaagg catgagccac attgcctgac 90840
cttgcatttt aactggagcc tgaagaagcc aaacttattc ccaaatgaac atttccacaa 90900
gtaactcaga tatcagtaac atttgtagaa atttactttg cacttttatt gaatcaatgt 90960
aatttttttt ttagttggga ggatgggaaa atcttggaac tccgaatttc acccaaatat 91020
ttaccatgtt taccaattta ccataaatga cacctctagt tttctgttca acaatatatt 91080
caaacagsct actaaatgac aaataatgtt acttgaaaaa tatcttcatg taataaarat 91140
aaacatatgg atatagatat cctgtgaatc aaaatttacc aacatgraaa atatgtgttg 91200
actttgttat ctagtttatc tttagtaatt ccaatacagt gcgaagcaag ttaacattat 91260
tgaagcagtg aaacatttgg agttttattt tggttatatt tcttgcaata agaaacatag 91320
actttaaaaa actcactgat gtctacactg attcctaatt acccagaaat ttcagactga 91380
acaccaatgt catggaaaaa ccccttttgca gaaggacatc tggtttaatc catcattgag 91440
gcttaattat ttttctgtgt gttatttctc acacagagaa caatgtatgt atagctatta 91500
caaaactgat tcagcactct ttctaatgct ttaatgtgaa ttacttggca atctatattg 91560
gaatacaagt tagaaaataa gagtcatgaa agataaattt tgctactttt tacacaaata 91620
ttaaataagt gtgcactttt tctaagcaca gggcacactg cgcatggtaa acattaaata 91680
agttatcaaa ggcaattcct acaggcatta aacatcacct ttgatacatt tcagcaaccc 91740
agtttaacca ttggcaaata aataacattt tgctaagtta cacattcacc aggctttcca 91800
taaaatatta cttcccagca cctggattaa ataccatgaa ctaactctac tacaagaatt 91860
ttgaggtcca aattatttta cccagtagaa gtttaaataa agaactagaa atgatcttca 91920
gatgacgtgt tattaaacat ttcacaatca ttcagctatt tattttcagt agactaaagg 91980
ctttagctct gatgcagaaa atttgaatat atcatgggag ttatcaaaat tattttaaat 92040
ttaccaatag tatttgtcat atgcatgctt ttagtaatct cagttgagca tatgcccccc 92100
caacatatat atacacatac atgaatatat atagaggtgc ttgtatgtta cactgaggcc 92160
ctctagaagt gaaacttgaa tgttttaata tgactaagtt cataatctcc tgcccaggaa 92220
attataaatt aaaaacattc cagatctcaa agagcttttc tttactgaaa tattgagaaa 92280
gacatttccc aagtacttta gcctttaaaa tatgattcat gcaagcaact aaattagaga 92340
gttcttggga agagggctgt ggataacaag agaaaagtga gaagagaaga aactaaggat 92400
gtgtgatcaa agaacaggtt aacatgtaaa aaattcaaaa gaggaaaaaa atgcttgcaa 92460
gttatatgag agtattttaa atatttggac gttgtctcat ggattacatc tatttatcta 92520
tctatctata tctatctatc tatctatcta tctgtctatc tatctctcca tctatctatc 92580
atctatctgt ctgtctatta gtttgactta tgccttacaa ccaataaaca cttaataagc 92640
aaccatgatt atttgttaaa tatgatatga tgaataaaca aatgctagta tttaggattt 92700
cacaatacag attgtttgag agtagaatgg attcacgtcc acctgcaatt gatacgcaga 92760
gctctgatat cacattctgt tttattttcc tcagagttta acatttttt ctcctgtttc 92820
tattctctcc ccatttatgg agtgcctctt cagtgctcag tatttacccc aagcattttt 92880
tttcctataa tcaatgatag ttctctacta gtggcacggc tgaatgaagc tggtttccca 92940
aattgtggag attatgtctt acttccttga attttgttta gtttttgagg ttcagggatg 93000
acagcattcc ctgctgctca gtgtgaaatt agggatttta gcctaaagag aaaatttgat 93060
agtgctcttg ggtctttgaa taaggaaggt tgggacatgc tgaaagcctt tgcccacttt 93120
ctggagcttt actttacaca tgtccatata gaaggatctg tgatgttctg caatagagca 93180
acctatttta cctttcccca cttaaaccca gtatttccaa actttttttt ccccataggc 93240
ctattttgtt gcttaatatt cagtggttct tctgaacaaa gttttgaaaa ttctaaccta 93300
aaaactcaga cgatagcaag attaacatgt ttacaacccc ttttccttac tcctaatggt 93360
acttgctcta catcagttga agggcttttc taaaagtgaa atgaaacaga agcggtgtga 93420
tttctttgta gacctaaagt gcctatgtca atgggaggag atactattat ccaaatatga 93480
gttttctctt taagctgtag agggtaggtt gctgcatata tttaaaggaa agagcaggac 93540
atagaaaatt aaattatagt atctttttta tccttttttc tttatttttg tcattatttt 93600
attatttgtt aagtataata tgatgattaa acagatgata atatacaaga tatattgcat 93660
aaatgtaaag attagataaa tatatttatg ccattaaaac ataagattta tatttactaa 93720
ggtgaatatt ttccaaaatg atgcatcctt ctggagagtg aggaaagggc acttttactc 93780
tttacataaa tctttatttt ctaattattg cattttcctc tcaaaggcca aagttgtaga 93840
tttaagacaa gaaatcattt caatttccaa gtaacaggga acgtcagttc taattcatca 93900
ccatcacttg ggtgccatac acacgcacag gcacacacac tcaaaaactg cgtgtcaaag 93960
acaaattcaa agatgtgctt ccttgtagat ttcaggaggc atgttcactc ttcttggtat 94020
gtgttggaga ctttaagttt tgtttgacta cttaccgctt attaaatgtc ttatttgcca 94080

caatagaaaa gtcatccctg catctatgtg aaatatgttt aatttctctt gaaagttgcc 94140
taagacaatt ttctgcatgt actgttcata ctagctaaaa ctgctcccca ctcttattcc 94200
tctaggaaat tcctagttat ttttcaagcc ccagttagat tattgtcctt tgatgcttac 94260
cctgattcct gaaacaatta gttattttgt ttgtattttt attattgaac taatcatatt 94320
taactttaat tttcatgctc cttaccatga aaatcaacca gctctttcaa ggcaagcact 94380
gtgatcagtt gtcttcaatt ccccagcaaa gcaacttgca tgcatggagt gttcagtgct 94440
gtttgtgcac aaatgtaacc atattacaat ggttaaatca tttagcatcc tgaaagcatc 94500
acagagtcaa agtagctaac ttgtgtgaac ccttaattca atttccggga agaggttaga 94560
cactgaagta tgtagttgaa gacactggtg ttgcaaggaa ttgtatccca ctagctcaag 94620
cctacagact tctgttttgt gccctggatt aagataaatt taaaaatcga atgaattgcc 94680
aacactatct ttggtagatt tcccataaaa acaatgaatt ctgatattta aaatgtaaag 94740
ctcttaagac cttacaacac aaggtccctg ctaacataaa gcaaaattct attggcctga 94800
gtgacttctg agtctacact gggaaccact ctatgtttca cttcagtgga ctcttaaacc 94860
caacccactt cactcattct tattagctaa agggctccca tagacattta agttttgatc 94920
tcttatctcc tttagtgttt tacttataaa ataacacaag gcaacattgg gaatttataa 94980
taaaaaatta ttgtatagaa tcccagcacc ataatttccc attgtctcat ttcattacaa 95040
ttgatattat ctgaattaat attatctgaa agtataatct gttcattata atttgtattg 95100
tttgaactct aggaataatt ctatctctct tgattacttt tcattttatt attctggaca 95160
ttataatatt ataaagaagt ttgaactaat gctgctagaa acctgaaaca aatttgtgtt 95220
ccactagcag aacaaatttg tcacattctg ttatatgctg taaacccctt aaatgcattc 95280
aggattaaat taaaacataa atggaccagt agataaaatc tatgaaataa catttggcaa 95340
cttctcttgc tttactccta caaactcctt cgactgaaaa ttccttcatt agtaattcat 95400
ataggccctt gttattccaa atttagaaac cttatctttg gtttattcac aaagaaatga 95460
attgcttttg attttgttgg tatagcagta agaaaaaaaa agtgaagaaa aaaaacccag 95520
aatttccttc aaacccctat gattctggac tctctaatac agcacatagc cagtgaaatt 95580
ctcaaaggtg gagacactaa agcaggggtc ccctaacccc caggccatga actggtatgg 95640
tccatgtcct gttagcaact gggccacaca gcaggaggca agccaccagc aagtgagcat 95700
tcctgcctga gctccgcctc ccgtcagatt agtcttggca ttagattctc atgggagcac 95760
ccaccctatt gtgaactgca tgcgcatgcc agctatctag gttgtgtgct ccttacgaga 95820
atctaatgct tgattatccc caacaacccc accctccaac ccccatccat ggaaaaattg 95880
tcttccacaa aactggtcgt tggtgctaaa aaggttgggg actgttgtta taaaggaagc 95940
ccatgattgt atagtcaact gaaaaaggca gttagtggct taatccattt ttcgacaaac 96000
tttaggacac tattccctca ttccaaaaca cgttggtaac acagaaacca ttctttttag 96060
gaatcatctg catatgatac acccagttaa ttggggcctt tgctttttta agaccaccag 96120
gccctatatt cttgagtcct gcctgactgg ctgtgtgagg gagaacaaga gagaaagaaa 96180
aagagagaga gcaaataaga gaaagagaaa aaaggtaaca tgtaatggac ttctcattca 96240
ccttgattta ttttcatcca tttaagtgaa aaatgctggt acctaagaaa atggttagag 96300
gaaattctac tttggtgacg gaatttattc tcttgggatt aaaggatctt ccagagcttc 96360
agcccatcct ctttgtactg ttcctgctaa tctacctgat cactgtcggg gggaaccttg 96420
ggatgttggt gttgatcagg atagattcac gcctccacac ccccatgtat ttctttcttg 96480
ctagtttgtc ctgcttggat ttgtattact ccactaatgt gactcccaag atgttggtga 96540
acttcttctc agacaagaaa gccatttcct atgctgcttg tttagtccag tgctatttct 96600
tcattgctgt ggtgattact gaatattata tgctagctgt aatggcctat gataggtatg 96660
tggccatctg taacccttg ctttacagca gcaagatgtc caaagggctc tgtattcgcc 96720
tgattgctgg tccatatgtc tatgggtttc ttagtggact gatggaaacc atgtggacat 96780
accacttgac cttctgtggc tccaatatca ttaatcactt ctactgtgct gacccacccc 96840
tcatccgact ttcctgctct gacactttca ttaaggaaac atccatgttt gtggtagcat 96900
gatttaacct ctccagctcc ctcatcataa tcctcatctc ctacatcttc attctcattg 96960
ccatcctgag gatgcgttct gctgaaagta ggcacaaagc gttctccacc tgcgggtccc 97020
acctggtggc agtgactgtg ttttatggaa ccctgttctg catgtacgtt agacctccca 97080
cggacaggtc agtggaacag tccaaagtca ttgctgtttt ctacactttt gtaagcccta 97140
tgttgaaccc catcatctat agtttgagga acaaggatgt gaaacaagct ttttggaaac 97200
tgatcagaag aaacgtgctt ttgaagtaaa atcagtgtat ctttattagt caaataaaaa 97260
aatctttcta tttatgagaa ctgtatttaa ctttagtagc tttacagtaa agtcaatttt 97320
agatccccaa tgaaaaacta taatctaaca acaaaacata gacatagata gagaaaattt 97380
atggttttac atgttacata tgagtgaaat atgtgtgtgt gaagggtagg gaggtgtata 97440
tgcatgtatg aggaataatt attcttttat tcataacatg catacattta catgtatgaa 97500
tttacatgtg taacaattta catgtaaatt tacatgtaaa ttacatttat aacaacatat 97560
tttttacatt ataaaaaatt atacattgtt gtaataaaac attaggtctc tatggaaagg 97620
taaaaagtaa gagattaaaa tattttattt tctttcccac accctgattt tttattctca 97680
cttcctagag ttacaacgat taacagtttc tttagttttt ttttttaat ttcttcgtat 97740
ctacaaacaa atgtgtggtt atatgtacac acatatatac acgattgtta aaacacaagt 97800
tgactgtcat ggttgctttt tagtggttta caattctgta gcaattacta ccaagaaact 97860

```
tcagtctcat ggtttagagg aatgtcacaa ctgctggtga gacttgctgt ccctctaggc  97920
tgtttctcta ggagagaaaa cttacgataa cacctctgtg tccatgcctg acacccactt  97980
atgtctctcc acttgtgaga gtcactttca ataagacacc tttctccttc tgccacgaca  98040
ggacatgagg ccagcccagc agttattttt caaatattaa ctctggaagc agacaaagaa  98100
tggctcctat aagggtatat gactggctgc aaataactgt gacttctgca gaacttatca  98160
attcaattct acacttctgg ggactagagt ccttcatgct ttctctcacc tctgcttttt  98220
taatctattt ttcaggatag atcatatatt tttattagta gcatcattca aattttactc  98280
acagtggtac tttgtcacag ttagttatta tatacctaat tttagaatga catatgtcga  98340
cttgtttgta agtattttag tgctaaagga agcaaaggac aagttgatat tttaaggatg  98400
tgacaaatct ggtgactttg cagttgataa cataagcaaa attgatgaat gtgtgctgag  98460
tatctacaat aagtcacact tctcctaaca ccaaggactt ttcaaggagc tcgacaatgg  98520
catagtttta caggggtcat gaaactgtaa acagaaagat caatacaatt tgttaaatgc  98580
tgtaagacta tggaggtgct aacgaaagag aggaaagaaa aaattatctc tgggcaaatt  98640
gaagctagcc ctgtgattgt ccaagactag aacttgttta tgctttatgc accttcctca  98700
gccctcccat agctcctagt atgacgtaat gtcaggtggg gtctgagcaa aataattaag  98760
aaccaatttg tgatcaaata tgtagatgtt agtttcaatg aaaaagaaaa aaccataaaa  98820
ctataacaag tcacaactca aaaagacatg atatttttc tcagttgtat ttcaaagctt  98880
tctttgtacc ttcgagttaa tctctgctta gtatttctac agaaattttc agactcattc  98940
attttatatt attgcatata attattttaa accttgtact actacatgtt cctatttaaa  99000
taccttgatt ttgaaagtca gattgcccac aaattctgtt aaaacacaca tataaataaa  99060
aaagctattt ttactaattt tttacattta attaacaatt gttgaattcc gcttagacac  99120
atacctagat gtattcctgt ttaaatacac atatgttttt agttatgtga aaccagcttc  99180
ttaattatat aaaatttttct gatctcaggg cccatgcttc tatctgtaca gcctagcttg  99240
gcaccgaaga aaggttcaat gactgcagca tgaaagagta cacatcatac acaattttat  99300
taaaagttaa tgaatgtcca ctttcaaagt aaacaaagtg cctcaatttt agaatttata  99360
agaaaagtgt ctttgaacag accagcagat gactctgagg gaatatgtta ttagtcataa  99420
gccaacttac ttttcagaaa tacaagtaac cataatgtca ttgatgttta tgactgactt  99480
tagcaataca aaaccaatag ttacgcccat atacacaaaa gttcaagata gaaaggaatt  99540
acaccaagtg ctttcatgaa ttgactggct cgctggaatc acgtctatca gcctgttaag  99600
actatctaag tccatatcat attgtattag acataaaata aatatgtgct gatgaaaatt  99660
tttggtatta atatgttatt ttagtaggcc ctggatttta gctaaaaagt aaagaataac  99720
tttcttattg atgacattaa aatcccatta ttaaagtctg tcataagaga tttcaagaag  99780
agtctccaat ttcatgtctc ttttctttca aagttaattc tgctgaattc tttatttgtt  99840
ggacaaattg ggtaatacaa tgaagaaata atcagaaaaa ataaagaata tttactgtac  99900
ttttgaaaag aacaataaat atgatatatt attttctcct cacctatcca aaagatcatc  99960
actctgatat tcctttcagg aaaatttata catgtgtata aaatactaaa tgttataggg 100020
agtagtgatg ttatgagcct aatgtcttta ctcatataac agatgttttt gggaacaatg 100080
aataagtgaa aaatatttag acactcacac tcacacatac acacaaaaat aagacagagt 100140
aatacttttt attatagttc aggtaatatt ttaatgtaag ttgttttact taatctgccc 100200
atcaacacta tgcattcagt actataatct ccactttaaa gaaaagacaa cagatttaca 100260
agttgttaac ctaaaagcac ataggtagat tagaagggga gctcaaatat gcctggcttt 100320
aaagtttatt tgtttgtttg tgttactatt atacttttgc cactatactt cattgaaggc 100380
aaaatcatac aatgaaaaat ataattattt tctatcttcc agaagtattg ctgatcatgt 100440
atttgacagc tgcatgcaca ctgaaggata taatgcacta taaaaaattc tttttataat 100500
gtatagataa atttaacttg ttccataaga caaatttttg cagtcagatc ctaactccat 100560
catttattag ctaaatgagt agttatgagc atgaactctg gaatctgatt tactgagcaa 100620
gttatacaac ctctctcggc ctcagtgccc tcttctataa agtgaagaca tagtgatttt 100680
tcataaggat gctgtcttca cagaagatat taattaacac atgaaacaca cttaggacag 100740
tactggatac atagtattta ctcaataaaa ggtacctgtt ttaattatta actggaaagt 100800
tactttacct atcttagact taggttcctt tcataacctt ctatgttttc ttttaataca 100860
tttattactc tttggatata attaatgtct gtgttctcta taatattgta agtgacatga 100920
ggtcagaaac aatgatgttt tgagagcaat tattttctat ttttcagcct ggtatgtagt 100980
atgtgatccc taaaaaatga tagatagatg attgatgggt agatgacaga tagatgatag 101040
atagatagat agatagaaaa aaagaagtaa aaaaaataat ggagattata atattagttt 101100
tgtaaggtaa tgggattttc ctgcaaaaca aaacactgat ggtaacttaa atgtgtatat 101160
tttattcatg tataatcagt ctgcttttgt ctccttcaac agaattcatc tggaaatgtg 101220
cagaataaat caaaccagag agaagaaatt tatcttcctg aggttggcca gttgcccggc 101280
gttgcatctt gtcttctttc tcttggttct attaatttgt atttttcttt ttctggaaaa 101340
cttggcccta aatctgcaca tcacatcctc acactgtctc cttgccaatc tctcaatcct 101400
acaaattctg gttaactttt tatttggcaa gttcaattcc atagccaatt tatatgaaca 101460
gcataacatg gtcacattcc cgttaaccca gtgttttgtc cctgcagcca tctgtggcaa 101520
tcacatccac atggtcatct gccttccatg tagtcacaat attgggtcca aagtgctcca 101580
cgtttatatg gctgtccttc tctatgtttg catttcatgt aaacttattt aacatggtgt 101640
```

tatccccaga gctgacgaca tggaatagca tcatatcctg ccaccttgtg gctactccga 101700
ccaaataaaa ccaaatttca ccacacgcca tgttgctact ccataatgtg atggaagtgt 101760
atgagagttc taactttata ttttccaaac aaatttggaa tctttctact ctttgtaagc 101820
cagaaaaact tctctgtttt tgaatttcac cttatgaata tgtctagata atgttacctc 101880
cttgaaggaa tcaataacac aaaacagtag ctgccctata tgtttgtgac tctttttgaa 101940
caaccaaata aatagaaata gaaagaacgt aacgtttgaa aatagtagca tgatggttaa 102000
ttttatgtgt caacttgact gggcgacagg gtgcacaaat acttggtaaa atgttatttc 102060
tgggtgtgtc tgtgaatgtg tgtctggaag agattttacc atttgaattg gaagactgag 102120
taaaggagag tgttctcacc aatgtgggct ggcatcatgc aatctgctga gggcctgaat 102180
agaatagaaa ggcaaaagaa gagcagattt gttttctctg ggtgagttgg gacatcaatc 102240
ttcgcctgct cttggacaat gttggtgctc ctggacctcg gacctgcaga ctctgaccag 102300
gacttaaact atttgttccc ctaccaggtc ttcaaactta gactgaatta aaccaccagc 102360
tttcctggtt ttccagcctg caggcggcaa atcacaggac tttttggcct ttatagttgc 102420
atgagccagt ttccattgtg tgtgtgcgtg tgtgtgtgtg taatcatgta atgttatata 102480
tatatatata tatatatata tatgtaaaat cttcattcta tatatagaat gcttctcttt 102540
ctctggagaa ccctgatgaa tacagatttt ggtattgaga gtggttctag agcaacataa 102600
ttttaaggat atattttctt atctgtttct gaggtttctg gaattggctc tttaatttca 102660
ttagatttaa agatgctaat gactatttag agtagtcgaa aaagcattga tagtccatga 102720
cataaactgt ttatacatat atgcaaagta tccacattgg attatcctca tcagacactt 102780
ataagatata aggaactatg tgactctcta tatacttttg tactttctta gaaaactaag 102840
gattataacg atattgattg gttactctta atgtcactgg acaaagtgct gaaagataag 102900
gatgagcttg gggattttaa tttccagtcc aaccctgaga gcttctatgt atgtctggaa 102960
ggagatcctt gtctcctgta gcttcaatta gagaattgca gaaagggatc ctatatgttg 103020
cgttgggggt ttgtgggaag atcctgatga agctggaaag actgagcccc taaattctga 103080
taagtgctat ctgacagtga aagaggtttc cctactccta tactcctagt ggaattggcc 103140
tccccgttcc cagtggtatc agcttttcca cctatgtctt gggtaattaa ctccatattg 103200
ccagaataaa tggtaaagac ttcccctgag gcagtttcca tacaagacaa tacagattct 103260
ttttggggcc caccttacc accttcttt gcttctaaac ttataactaa acccaagtac 103320
caacagcccc tgaaagaaga ggtaccaagc aagaccatga ggtggtgcac taaaaaagta 103380
gatgagtttt ctaatttata caaacagaaa tccagggaac atgtgtagga atggatacta 103440
agcatatggt aaaagagtag aatgaagata tagttggatc aggtctaatt tgtcaatatg 103500
agctcattaa acagagattc tccatttaat gttgcaactt ggggagttag aaaaagctct 103560
aagtttggtt gtttggctga aacatgaatc aaaagatggt ccactgtgag taaactggaa 103620
atgtttaact tcccttggtt taatgtagag gaaggaattc aaaggtctag agaaattaaa 103680
attcaagagt ggacttgcca ttgaagacct actcacccat actggaaggg ttcagaagac 103740
aagcttttca ccaatacatt gataaataaa cttgtgaggg gaaactgcaa ccacacaatt 103800
ggaaaatcaa aatgcagtaa tagtaactag atcatggggt ggcagaggcc aggtggtggc 103860
tttcaatggc ccaaggcaag gtgggcatag ttaccataat gaacagcagg ttaaagcatc 103920
aaatagaata gtctgaccca cacgtatcta tgacattggc tagcacaggc tagttaatta 103980
tggtgttccc agttgtataa gtaaatagaa agcccactaa attcttactt gatctgaata 104040
tgcaaaaaca ttttaggtca agtgaacaaa agtctattct gaattataat aactgatagt 104100
tacagtccct caatcaattt gtagacttga ggcagtttac agacccagaa ctccttaaat 104160
gaaggggaga caaggtcttc tcaaggaaga tttccatata ttgctaggtc ccacagagag 104220
ggacctttta tctgggtaac tgcattggag aaaaggagat aatcagactt ttgagtacta 104280
cttgacactg gctctgaact gacaatttca ggctcacaac atcattatag ccccttcagt 104340
cagtaggggc ttatgatgat tagattatca gtgaattcca tctcaaagtg ggcccagtgg 104400
gtccctgaac tcagcctgtg gttatttctc tggttctaaa atgcataaat agaatagaca 104460
tactggcaga atccccatgt tggtcctcct acttgtggag cttgggtaaa agaaatagtt 104520
gaaataattt atttaaagtg ttttgtgagc acaaagaaag aaaaaaatta tcttgacaaa 104580
aagctttttg ggtaattctc ccagagaagg taaccactga gatgggtttt gaaagttgag 104640
taggaatttc aaagggtaga ggatgaattg caaagacttt tcctaaagag tgaaccaact 104700
gatcaagaag aatagcttgg ttgtgcttat gacacaaata atagatagga aagtagtaaa 104760
taataaagtt gaatagtaaa ctagatcaga taaggcaaga atgaaatgct tttttaacaa 104820
atttagactt aattggatct acatgatgtg gtgatttgaa cgtatcctca aaaataatgt 104880
gttggaaagt taatccccac cgtaacagtg ttaaaaagta gggcctaatg ggatgtgatt 104940
aggtcacaaa agttacaacc ttgtgggatt gataaaaagg ctaaataaag gattactttt 105000
tttaaaaaag caattataaa tgggcttgag gctgcaagtt ctctaacttg ctgtcccctt 105060
ttcttgccct ctgccttcct ccatgggctg acacagcatg aaggccttta gatagatgtc 105120
agcaccatgc attttgactt ctcggttacc agaactataa aacaggtaca tttctattta 105180
ctataaatta cttagtctgt gttattctgt tgtagcagca taaaacagac taagagaaat 105240
ttggtactca gaagtgtggc tgttctatat cagctgaaca tgtgaaagtg gccttagaac 105300
tgagtaatga gtagaggcca aaagaatttg caggaacagt ctagaaaaag cctagaatgc 105360
catgagtgga ggtttaaggg tgattatggt gagggccagg aaatgtaggg aaagtctgga 105420

acttcttaaa gactacatgt gtggctatca tcaaagtgct ggtagaaata ggggtagcaa 105480
aggccattct gaaaagatct caggtgaaac tgagaaacaa tgtactggaa actagagtaa 105540
aggccatctt ttttatacag tggcaaataa attggcagaa ttgtgtccat gtagcagggc 105600
tctgtggaat acagaactta agagtggtga gctagactat ctggaagaag aaatatctgc 105660
gaagcaaaaa attaaaggtg gtgcacagct tattttgggt gcttacagta aaatgagaga 105720
agaaagaaaa aaaagataga atttgtaatg aaaaggaagc agcatgaagc gatttagaaa 105780
actcttagcc tggccatgta aataataaaa aagcatgttt gggacagaat actaaggaag 105840
tgttcatgta ggatacagta aattcctctt caaagtttag cctgttaact tcaagaagga 105900
gaaaaattgt taagtacaaa gagttctgag ttccacttca aagaaccaat caatatgtca 105960
gtatgtttag cttccctgtt ctttgttctc cattttaaag tttaacctcc tcgttcttta 106020
catctccttg cccctagttt cagtaaacaa ccccctccta gcctctatca cctgttctgt 106080
ccttagtcat cctcagtcac ctgctctgtc cttagttatc tagtcatcta cactgtaacc 106140
gtccttcccg ccgaaactac tcaccctgcc actctggctc ataccccctgc tctctttaaa 106200
atagccagtc agaattagct tagactgtgt ggtccaaccc cagccaatag gggaaagaca 106260
gcagtaggga ctagctgctt taggataagc acctcttccc ctcccttgta cggtgtgctc 106320
tcgccatcgc tccatccttg agatgcaccc ttctatagaa gtaaattgcc ttggtgagaa 106380
aacttttgcc tgagtgctag tttcactttg tagcaccgaa catttacttc caacaatcat 106440
ttgctacaga gatgaacata aatagaagga atccaggttg gatttatcaa gacaatggga 106500
ggatgaccct gaaagcattt caaagatctt gtggaaaagc tgggattttt agggcaaaga 106560
atccagagag atacataaat atatgtttag tttttaactg gaagtgagct gggatgtact 106620
gtatgtagta caatcatata ttttaacgat gtggaaactg aaatagggtg actagcctaa 106680
aactaatgga ctagcttgtg aaaacaggat ttctcgtctc aaaaaaaaaa aaaaaaaaga 106740
aaacagggtt tctcagtaga actaaactgc ctttccttgg ttttatttct agttaaatgt 106800
cgtaataaat ttaccagata ttgcttccgt gtacctggaa agcatagaca ttttgattca 106860
ttgtaatcaa tgtttaatca tgaaattaga tgaaataaca gcaagtgaag catgttagtg 106920
catgttagtg gctagagaga ggagattcat agtcaaagct gctttcttct ctcagtgagc 106980
tcactcaaca gctttttgaa tttgagaaaa ttgaattttc tttctaaggc tcagttcctt 107040
caatatataa tgggtctaat aatatttatt tcatattttt aatatttaca taaaattata 107100
cataaaaagt tctgtacatt gaatttataa aaacttttta ttttaagtat acttaattac 107160
cccaaatttt aacatctgta tgtggtacat catggcctag gaaaataaag aataataaaa 107220
attatcctaa tcatttgcac tctagatttc ttcaattata ttcatttact cttggacttt 107280
tctacaaaca aagcaacaca aaataaacaa agaaaaagaa ggggagcaaa aataccctat 107340
gacttttttt ttttttcaa atgaatgtat gctcaccgta atgataccgt gtggtagaac 107400
atgtttttgc tgcattagtg gccaagtgaa aagagtgagt aaggtgttcc tttttgagtg 107460
cagttctttc tgccaagtaa tggagggagt gaggtagaac gatgcccagt gatacacgat 107520
aagcaataag taaagtcatt accaagagat gtgaaaaaga ggacacagaa gagatgttaa 107580
gccattttaa aatccatgtt caacatcact gagctgtctt cagggaacat ttaattggca 107640
gaaatgactg cttctttgtg atttcttttt aatgcaacag ggaaaaggag tcaactttat 107700
aaacacaagc aaacctttcc ttatttatta tctctgaagg cccattggtt ttcctgcttc 107760
tttataggta gggctgtcta tttgtgtgtg tgtgttagga actagacaaa gtgttgctat 107820
ttatatttag aagccaaata cttattcagt agaaatgatg catgattaga aatagttaaa 107880
taacttcata ggagtatggg tggttatata ggtatcaatg cacacaatct gagaatttgt 107940
atttgttcta actactgaac aaatgggact aagaaagtaa aatttcttgc aatggcttta 108000
tgtgcctcac atgttgtcag atctttacat agattataac ttttaatcct caaaagatat 108060
tccttcaaca acattttgaa gtgagaaaac taaagctcaa aaaagttatg taagctgtct 108120
acaattgtag agctaacatg cagcatagtt gtaatttttt aagtcaaata aggcaagttc 108180
caaaactttt gagttggttt tgggagatga ataccagtac taaggatgtt ggaaacattt 108240
tagcacaaaa taattttaat gttttggaaa gtacaaatgt acatggcaga ctatttctaa 108300
ggaaggtggg atgtaaggtt aggacatgaa tatatagaag agtgcaagat gatgttccag 108360
tgttctccag aaaaaaagat gaagaaaaga ttgcaaagca tctaactacc caatcctcat 108420
gaaaacatat gcatgcacac tgaaataatt aaaaactaca tgtgtatgca tgcatgcatg 108480
tacacatgtg tgtattatat atgtctatgt gtatacatac aatatacata gtcagggaga 108540
ctttatatac atagtcaggg gaagagagag agaggttaaa tcaccatcct tctaaagata 108600
atagttccta ttacatgaac agccttttag ttcgtcaaag caccttaatg tgttcactag 108660
gcaaagttgg agctagagta gaatttgaga ttaggagacc ccaaaaggtg aagaacttag 108720
aataaaaata tagagaaaaa tacagttttt gaggtcattt tagaagagaa ggaaggagag 108780
gtgagaaata ggaggggagt agaattaaga gaaaccaggg aggtatcatg aattgggtta 108840
agtgctgtaa aataaccggt gacaaatcct ctatgattag gcaacaggta gaaaatttga 108900
gaaatataat agatttaatg ttttaatttt aatttttatt ctcttttgca acaaatatct 108960
atcttctact tgccttccac agagttgcaa tattatctgt gtctgtgtgt ttgcatactt 109020
gtaaacatta cagatttcta aatggcgtgg aatagatagg agttttaaga tgattttgaa 109080
aggtctcaag ttgctgacat caggattatc taagcattat ccttacattt tataagtcat 109140
accatagtcc ttgaatataa tctctcacag attaatttaa ttctaactaa taacaatcac 109200

ttcatatgtt ttcatcttat atagcagtca agtttacaat aatcttatta agcaccaacc 109260
ttaagagtat ttatattagt tagataacgg attaaattcc tatgagaaca aaacaaaata 109320
acagttaaac aaattagaag tttatttctc atttatgtaa aattctgagc tagtatgtca 109380
tatcagctct gccaaggttt taggagctcc aactaattta ctcttctttc tctgctttct 109440
ataaagagtt actctcatct atatgattca ggatggctca gcactatatc cacattacag 109500
ccagtgagag gcagaaaaga agaagggaag ggcaaacctg ttttatcaat ggatggataa 109560
aagtagctta attttctttt attacattct cttggccaga agctaccctt ttgccatagc 109620
aaatctaatt ttagaatttt aaaaaaagga aattctggtc atttttgttt tgttttcttt 109680
ggctttgttt tttaaaattt atttaactta tttatttta agagctgtcc cataattatc 109740
agttctggct ctgaggagca aatacttaac ctggttaagg gaaaggtact agcttctagt 109800
gcaggctcag gctctgtcaa atctgtctgt tgtctggatg tctggatgat cacttagcca 109860
tctaatcact tcgctatgca tgctcagtat gcacaaatat agacagacat gttggacaaa 109920
gtcatttcca tgctctaatt caggtctgtc gctccaaaaa attacctact ctgaagttta 109980
tgaaattaac tagaagcatg ttagtagaca ttttgaaaat ttttctgtaa aattttcaat 110040
tggtatcctc ctcaaatatt tatcttttat tttcccccaa aagtagcagc tagtctaaat 110100
tagcatattc tgtttgtcac ataaagaatg agctaaaatg tatttttagt tttttactac 110160
ctatccttat aacaggaacc taagggattt tatctgacat ggcatggaag aattaactgt 110220
aatcagaaaa agaatgtcag gtaccaggtt ctcaacttac agttaagtta gataacttat 110280
agctaagtta tctaaacaat taatttacaa aagtgctgag tattcatttg tttattagca 110340
ctagtggcaa aatctgagca ataaagatgt gagataaaaa agcaaaatgt aaaaaaaaat 110400
agacacagta gaaatgaata tgagtaagaa attaagtgac ataaaccgaa aagattagac 110460
cctgtttcaa ttaaatcact atctctagtg tctcctgaca caaaacaaat ttagaaaaac 110520
tgcagaggag caatgcaaaa caatgtttca agaaaatata aaaatgtttg ccctttagaa 110580
aatatctact aatccaacat tctcatttcc cactgaagga aattatgaga agaaactgta 110640
cattggtgac tgagttcatt ctcctgggac tggccaatca ccgggaatta cagattttcc 110700
tcttcacgct gtttctcacc atttacatgg tcacggtggc aggaaatctt ggcatgattg 110760
ccctcatcca ggccaacgcc ccggctccac acgcccatgt acttttttcct gagcaactta 110820
tcctttgtgg atctgtgctt ctcttccaat gtgactccaa ggatgctgga gattttcctt 110880
tcagagaaga aaagcatttc ctatcctgcc cgtcttgtgc agtgttacct ttttatcacc 110940
ttggtccacg ttgagctcta catcctggct gtgatggcct ttgaccggta catggccatc 111000
tgcaaccctc tgctttatgg cagcagaatg tccaagagcg tgtgctcttt cctcatcaca 111060
gtgctttatg tgtatggagc actcactggc ctgatggaga ctatgtggac ctacaaccta 111120
gccttctgtg gccccagtga aattaatcac ttctactgtg tggacccacc actgattaag 111180
ctggcttgtt ctgacaccta caacaaggag gtgtcaatgt ttgttgtggc tggtttcaac 111240
ttcacttatc ctctccttat catcctcatt tcctatctct acatatttcc tgccacccta 111300
aggatctgct ctacagaagg caggcacaaa gctttttcta cctgtggctc ccatctgaca 111360
gccgttacta ttttctattc agctcttttc ttcatgtatc tcagacgtcc atcagaagag 111420
tccatggagc aggggaaaat ggtagctgta ttttatacca ctgtaatccc catgttgaat 111480
cccatgatct acagtctgag gaacaaagat gtgaaagagg cattatgcaa agaactgttc 111540
aaaagaaaat tgttttctaa ataaacatta ctactgattt ttgttgtgtt gtcattttat 111600
ttaccctatg atttttcata gagcatagtc tgatacaaat atatccaaaa ttatatattt 111660
tcctagaagt gttggggaat ttttatgaag tgtgaataca agaaaaaatg agtatttaca 111720
tcaattaaca ggaattttga attcaatata gttttgcctg cccatacccc aaagtaagaa 111780
gctctatagt gaagaatcaa tgtaaatttt aaaaaatcta atttatactt tagagaaaaa 111840
aaatccttaa aaccagacct gggttttctt tttgcaagga gttaggcata tagttttaga 111900
ctgcttctta aacatatatt aagctttttt taaaaaaaac aacaacaaca acagcaacaa 111960
caaaaagaaa acagcatctc actccatcac ccaggatgga cctgagtgca gcggtgtgat 112020
cacagcttac tgcaatcttg atctctggtg ctcaaatcat cctcccacct cagcccctca 112080
cctctgccat agtagttggg actacaggtg tacaccatga tacctgacta atttttgtgt 112140
tttttttttgt agagacggga tttcaccatg ctgcccagac tggtcttgaa ttgctgagct 112200
ccagtaatct acctgcctgg gcttcccaaa atttaggaat tacaggtgta agccactctg 112260
cccagcctat aataagcttt aagaaaaata gcagttctgc taaactctac taaatacttt 112320
tgtgtcttac aaagaaagtc aacattccac tttaccatat tcaacctaaa atatgagcct 112380
catcccttgt acactcagct gtcatctcag ctgttgcctc tgcctgtacc aacctcatct 112440
ccagccaaaa acacaccaac actttggtcc acagtgaggc acacttatgt ttccacagtc 112500
catacctggt ttatttcttg gctttttaga ggaaaaataa atggctagta gacttctaaa 112560
ttgaatgtta cctaaaagtc tgtctaagga gatccaaggc taacagacaa tctgttcaat 112620
gcatgtctaa aaccacatca acccaaaagg tgtaacatgt ggttgtaatg atactatcat 112680
cacagatttg catgggatca ctgtcttcca gtcttgaaga tgtgtgtttt tgggtgtgga 112740
ggcctatagt tcccctagag gtaacctaga acttaagata taatcacata caaggaagcg 112800
agagactaat gatgaggtgt gtgcaaggtg agaaattaga acacagatgc cctcatgaaa 112860
ttcggattat tgaagggcta caacctctag agtaggacag agtcaacaac agcaacaaca 112920
aggatgaaac tagtccgttt agacttggtt tcatggagga gttaaaaagt cttatttaag 112980

aatttatgac aataattaat gctcatgtta agtcagcatt tatatttata ctgctttgat 113040
gggctatcaa atctcaagct aaataatgaa gaacataaag aggcccaggt tagaagaagc 113100
ttcatgacac tggaataagc aaacataaac ataagtcaga gaatgggtta tcaaccactg 113160
agacttcaaa gatagcccac tgaaaaagtt ctaattcaca gaaattttcc aggaaaataa 113220
gaaataaaaa gcacaaatga atatctttaa tttatttcag atattaggat tacttgatgc 113280
aaattgtaaa ataaagatat ttaatcattg aaaataaatg tgaaaattta aactctaaga 113340
ccatgtatat atgtgtgtac atatgtatgt gtacatatat atatgcaaag actgattata 113400
gaaaccaata attatatata atatgtagag ttccaagcat atctgcaggc tgaaggagag 113460
tctctgtgtt ttataggcat tgaagcattt ttagagtgga taggcaggta tataaatttt 113520
cacattgaaa ggaagaggaa gaaacagcta agtccctctc atgaggcttt aagtggtctt 113580
ggagagtcag cctttctctt cttcttcctt agtaaatttc atgcagacct tggcagccta 113640
atcactacaa tcctgcatca gtgagtcctc tgaactggca ccatagttca ctaggaaagg 113700
cactcatcca ccaagttcag agaggtggcc atccgcattg ctaaacagaa taaaacaaaa 113760
attcactttg tcatctgcct gacattggcc ttgctataag atagcctcat aaactgcctg 113820
aaaggcctta ttcttttcaa aggattttgt gctgtcttca agtgatattt ttctccatgt 113880
cagaaataaa atatttcagg attgacctgc ctcaaatttt agtttgtctt ggttattggc 113940
cattgcatgt ataagtctaa ttgtgtgaca gaagtttccg atactctgct tcatgaagta 114000
ccctttagga ctatctcctt gtctcttcag ttcttcaatc aatttttcct gaagttcttg 114060
tgtgaggctc tgatacaaag agtaggcagg tgctggcatc cagcccagag ccacccttte 114120
cagccctagc acatccttaa agcacccctg accagcaccc tcaacctggc caacaggttt 114180
ttcagcatct acaggttaac ctccatgggc ttgagctgca ccttcacaaa tcacaagaga 114240
aaccacaaca cagagtacag ggaaaacagc ttttcaaaat tggtaaggtt agtagtggtg 114300
ttaagatatc accctgtctt cctatttgaa taccctattt ttttttgcc tgattgccct 114360
ggccagatct tccaatactg tgtttaatag gagtggtgag agagggcatc cttgtcttgt 114420
gtcacttttc aaagggaatg tttccagctt ttgcccattc agtatgatat tggctgtgag 114480
tttgtcataa atagctctta ttatttgag atacattcca tcaataccta gtttattgag 114540
agttttcagc atgaaggggt gttgaatttc atagaacgcc tttttttctt catctattga 114600
gataatcatg tggattttgt cattcgttct gtttatgtga tggattatgt ttattgactt 114660
gggtatgtta agttagcctt gcatcccagc gataaagcca acttgatcgt ggtggataag 114720
ctttttgatg tgctgttgga tttggtttgc cagtatttta ttgaggatat ttgcatcgat 114780
gttcatcagg gatactggcc tgaaattttc tttttttgtt gtgtctctgc caggttttga 114840
tatcaggatg acgctggcct ataaaatga gttagggagg agtccctctt tttctgttat 114900
ttggaatagt ttcagaagga atggtaccag ctcctgtttg tacctctggt agaattcggc 114960
tgtgattcca tctggtcctg ggtttttttt gattggtagg ctattaatta ttgccacaat 115020
ttcagaactt gttattgatc tatccaggga tttggcttct tccttgtttt ggagatgaca 115080
tgactgtata tttagaaaac cccattgtct cagccccaaa tctccttaag ctgatgaaca 115140
tcttcagcaa agtcttagga tacaaaatca atgtgcaaaa atcacaagca ttcctataca 115200
ccaataacag acaaacagag agacaaatca tgagtgaatt cccattcaca attgctacaa 115260
agagaataaa atgcctagga atccaactta caagggatgt gaaggacctc ttcaaggaga 115320
actacaaacc accactcaag gaaataagag aggacacaaa cagatggaaa aacattccat 115380
gctcatggat aggaagaatc aatatcatga aaatggtcaa actgcccaaa gtaatttatt 115440
gattcaatgc tatccctgtc aagctaccat tgactttctt cacagaatta gaaaaatctc 115500
ttgtaaattt catatgaaac tgaaaaagaa cccatatagc caagacaatc ctaagcaaaa 115560
agaactaagc tggaagcatc atactacctg acttcaaact atactacaag gctacagtaa 115620
caaaaacagc atgatactgg taccaaaaca catatataga ccaatggaac agaacagagg 115680
cctcagaaat gatgccacac atctacaacc atctgatctt tgacaaacct ggcaaaagca 115740
agcaatagga taaaggattc cctgtttaat aaatgatgtt gggaaaactg gctagccata 115800
tgcagaaaac tgaagctgga ccccttcctt atgccttaga caaaaattaa ctcaagaggg 115860
ataaaagact taaacgtaag acctaaagcc acaaaaaccc tagaagaaaa cctaggcaat 115920
accattcagg acataggaac aggcaaagac ttcatgacta aaacatcaaa agcaatggca 115980
acagaagcca aaactggcaa atgggatctc actaaactaa agagcttctg cacagcaaag 116040
gaaactatca tcagagtgaa caggcaacct acaaaatggg agaaaatttt tgcaatctat 116100
ccatccagaa tctataagga acttaaacaa atttacagga aataaacaaa cgaccccatc 116160
aaaaagtggg caaaagatat gaacagacac ttcacagaag aagacattta tgtgggcaat 116220
aaacatgtga aaaaaacctc atcatcactg gtcattagag aaatgcacat caaaaaaaca 116280
caatgagata ccatctcacg ccagttagaa tggcgatcat gaaaaagtca ggaaacaaca 116340
gatgctggag aggacgtgga gaaacaggaa tgcttttaca ttgttagtgg gagtataaat 116400
tagttcaacc attgtggaag acagtgtggt gattcctcga ggatctagaa ccggaaatac 116460
catttgaccc aacaatccca ttactgggtt acatacccaa aagattataa atattctact 116520
ataaagacac atgcacatgt atgttcattg cagcactatt cacaatagca aagacttgaa 116580
accaacacaa atgcccatca accatagact ggataaagaa aatatggcac atatacacca 116640
tggcacacta tgcagacata aaagaggatg agttcatgtc ctatgcaggg aaatggatga 116700
acctggaaac catcattctc agcaaactaa cacaggaaca gaaaacaaaa cactgcatgt 116760

tgttctcact tataagtggg tgtgaacaat gagaacacat ggacacagga aggggaatat 116820
cacacaccag ggcttgtcag gggttcgggg gctagggaag ggatagcact aggagaaata 116880
cttaatgtag attacaggtt gatgggtgca gcaaaccacc atagcacata tatacctatg 116940
taacaaacct gcaggttctg cacatgtatc ccagaactta aagtgtaata aaacaaatgt 117000
atgtcacctt gttaaaaaac agatagtaga aaaaccagca aaatgagaac tggtgctata 117060
taaaaatgta ttcaaagaac caagtaacat atgactttgt gctcaacttt aataatcatt 117120
agggaagtac acattacaat tacaagatat ttgtgttttt atatatctgt acctatatgt 117180
tgctaaatga caaaaagaaa aaaatgatat tgctaagtac tggtgggact atggaacaat 117240
tgaatctctc acataatgtt agtgagaata caaatggata caaccagttt ggaaaactat 117300
taggcagcat aaatttaagc tggacatatt tatactccat aagccagcac ttctactcat 117360
agttacagca gtcagacatg catgaatatg ttcatgaaga gatatgcaca agactgaatt 117420
ttgtaatagc cacatctgga aataacctat atataaagtt tggtgtactt attatacagc 117480
agtgcagatg aaggaattat ttacatattc atcacactca tcatcatgaa taaaaatcac 117540
caacataata tttagcacca atgtagtgtt cattggagac tagacctaac aaaatatgca 117600
ctgtataatt tctattacag aaaattcaaa accagacaca attaatctac agtgttaaaa 117660
gcaaagatat agttaatttg aggttactga ctaggaaagg gcatgaagac agatggtgga 117720
gtactggtaa tcttctattt gagaatttag aagcaagtaa tacaagtatg ttcactttgt 117780
aaaaattcat gaagcaggcc gggcacggtg gctcgtgcct gtaatcccag gactttggga 117840
ggccgaggca ggtgaatcac gaggtcagga ggttgagacc atcctggcca gcacagtgaa 117900
accctgtctc tactaaaaat acaaaaaatt agccaggcgt ggtggcgggc acccgtagtc 117960
ccagctacta gggaggctga ggcaggagaa tggcatgaac ccggaggcgg agcttgcagt 118020
gagccgagat cccaccactg cactccagcc tgggcaacag agcgacactc tgactcaaaa 118080
aaaaaaaaaa aatcataatt catgaagctg tacacttacg ttttgtgctc attatttgtg 118140
tgtaaattag acttcaatat aaagcttact aaaaacgaat aaaaatagta ctagtcttca 118200
agcaagcaaa gcttcattcc aatatcaaag cattctattt acctatcagt acacagaggg 118260
tattagtttg ctagggctgc cacaaataag taccatgaac ttggtgactt aaacatgcag 118320
atttatttcc tcacagttct agaggctaga agtccaagat caaggtgtgg gcaaaattgg 118380
tttcattctg agttctcttt ctggcttgtg gatgatcatc ttatcccgac ctctttacac 118440
tttatttttc tgtgtgtatc tgtattctaa tctcttctta taaggatgca agttatattg 118500
gattagggca cagctcaccc attaggcttt attttactaa atgttctctt tagatattgt 118560
gtctccaaca gtcatgctct gtggtgcttg gagttagaac ttcagcatat gaattttgga 118620
gagggaggga aggggcacaa tccagtccat aacacagatt aagaaacgtg aaaggctaat 118680
agaagtttga cacaaagttt gtgacactag tacgagagaa actgtatcag aaaagttgaa 118740
ttaagttgaa agtaacatgg taaacctaag gcaatgtgag aatccatggc agacatgaat 118800
gttatcttat ggattttcca atgtaagaag gaaaatactg agaatgaaac ataaggcaga 118860
gaaggaccag agagttgtga gttcccattt taaatttgtg ttgtgccaaa tgtcatatct 118920
ctagagaaat tattcagtga gaaaaaaaat ctgacagagt aattgcttca tttttgcata 118980
tctgtgaaat cccttaggga aataaagtca tcatacaaat attataaatt attcctgtat 119040
ttgtcaccag aaaagccatt tgatattctt tgtaaggata gctcttccct tattcataaa 119100
taagtttctg catgtgtttg taatcctgga acacttgttg tacaatcata tgtattttca 119160
gagttagata tatgattgtg atgattaaat gactaggtag aaagaaaaat gccaattacc 119220
agaaaaatgt agacacttag catttaaggt acttttattt gttaaagtct tgaataatga 119280
ggatggaagt taatggcata aaaatataag aggcatgctc taggatcttt cactcaatat 119340
aaatgaaagc taatatttat taagggtttg ccacacattg ggcacagtgc tatgcatatc 119400
acatacccca ttttgtgaaa tccgaaaaag agtgcttttg tatttgctaa tttcatctgt 119460
aatagaaaaa aattatagtc cagaattatt aagaaacatg acccagacta ctcagatcag 119520
aagttctgac atcagaatgt gaactcaacc agtcgactcc caaacgtatg tttctaccag 119580
tacagtatgc tttatggttt gtagtggaat ttctttctgt actaaccatg agggaaatat 119640
gttattatcc atatctatta taggcaaaat gtcatagaat tggtttgagg gttgaatgtg 119700
ttaaaactta taaaatagat tagtgtttgg cttataagaa acaccatgta agtgctggtt 119760
aaattagtgg taaaactaaa acacagaata aggaacatgt caaaagaaca gagcagcatt 119820
tcagaaatat ctaactccag atcctgtgaa ttgattttat gctaaggcta tcatattttt 119880
atcaaggcat tcatctttgt gtttggtcta gtcctaaaac ttaagaatgt caaccggatg 119940
tgtgcatttg tcaaatacac aaagttgggc actttaaata tatgaatttc actgtatata 120000
aattattgca taaacaacat taaacattaa acaaagaaac aaaggtgaaa tctgacagga 120060
gcttgatata taaaaatgaa tgaaggtatg tagggaaaag aaagagagat cagactgtta 120120
ctgtgtctgt gtagaaagga aagacataag agactccatt ttgaaaaaga cctgtacttt 120180
aaacaattgc tttgctgaga tgttgttaat ttgtagcttt gccccagcca ctttgaccca 120240
accactttga cccaacctgg agctcacaaa aacatgtgtt gtatgaaatc aaggtttaag 120300
tgatctaggg ctgtgcagga catgccttgt taacaaaatg tttacaagca gtatacattg 120360
gtaaaagtca tcgccattct ctagtcttga taaaccagga gcacaatgca ctgtggaaag 120420
ccgcagggac ctctgccctt gaaagcggag tattgtccaa ggtttctccc catgtgatag 120480
tctgaaatat ggtctgaaac caggggcaca ttgcactgcg gaaagccgta gggacctctg 120540

ccccttgaaag cggggtattg tccaaggttt ctccccatgt gatagtctga aatatggcct 120600
tgtggcatga gaaagacctg accatccccc agcccgacac ccgtaaaggg tctgtgatga 120660
ggaggattag taaaagagga aagcctcttg cagttgagat agaggaaggc cactgtctcc 120720
tgcctgcccc tggggactga atgtctcggt ataaaacccg atagtacatt tgttcaattc 120780
tgagatcaga taaaaactgc cctatggtgg gaggtgagac acgtttgcag caatgctgcc 120840
ttgttattct ttactccact gagatgtttg ggtggagaga aacataaatc tggcttatgt 120900
gcatgtccag tcatagtacc ttcccttgaa cttaattatg acatagattc tattgctcac 120960
atgtttgttg ctgaccttct tattatcacc ctgccctcct actacattcc tttttgctga 121020
aataatgaag ataataatta ataaaaactg agggaactca gaggctggtg caggtccttg 121080
atatgctgag tgccggtccc ctgggcccat tgttgtttct ctgtactttg tctctgtgtc 121140
ttatttcttt tctcagtctt tcatcccacc caactagaaa tacccacagg tgtggagggg 121200
caggccaccc cttcaaaggt acatcactac tatgattgaa tagatgtaga cacagctttt 121260
actcgatgtt gtataaatca caaataaaat ctttgtttca gatattcaaa aataaccttc 121320
tataagatgt tcttgtgaat atagatgatt tttatataga aaaatagaag gtacatttca 121380
ttaatatcag tgttaacagt ttttgttgtt gttgtaagta acagaacaca gctctggtta 121440
tgctaagcaa aaccggaaca ttctgaaagg atgttagggc tcctaaatca atgtgcggtt 121500
aaaagaccag gcttagagaa cagatagcag cctaggcagg tgtggaggct gagcagatca 121560
aaccacacag aatcagtgtt tctttgcaac atcagcctga tcttcaaacc tcactgtaat 121620
gaattagatg aaaaaaaatt ttgtcaaact tgagtcatta actcactccc ttaagacaag 121680
gagaggctag catctgaccc ctgcttaagc ctttctaaac agagaatctc atctctcagt 121740
ggaaactaga agtggaagat caagaggaca gaatttccat gactctcaac ttgacacatt 121800
ggaaaatttc ctccaaaaaa ttcaagtgga gtgttaatgg ctggaattta atagctgtgc 121860
aaccctcgaa acgcaaatat gtagatattt ttcaaaataa ggggaacaaa tcatgccctg 121920
cctggaaaat ctctattgca ttgagtctgg aggaagaaga aagcaacaag tgagccccgc 121980
ttgtttgctc aggcttctcc aagaggggat tgaactggtg aaatctattc agaaataaca 122040
agatcattca gtcaatattg ttctttgata tcttgttctt tcctctctct aaactcattc 122100
tctgccagaa ttataatcag gaataaagaa attggttagt tttaaatgtt atgttttggg 122160
gcaagtccac ctataaagta catttctcaa atataaagag gggcagaata agagaaggag 122220
aaagaagaga agaagaatag agacagaagc tgaattagaa gggaatgaaa agaaagacaa 122280
agggacataa aacaagaaaa ataaagggca gtatcaaaaa aggcagagag ctgctttgaa 122340
tatccctact ttagtttacc tgaaacaaag ttcattagaa ataaaatgag gaaaactgct 122400
gaaaagaggg gaaaactttt ggtaagaaga ttattactct actggtaata atagtaataa 122460
tcatgttgct attgtattgc aatcaatata tatgtaaagt tcctttctcc cactgaagga 122520
aattatgaga aggaacttca cgttggtgac tgagttcatt ctcctgggac tgacgaatca 122580
ccaggaatta cagattctcc tcttcatgct gtttctggcc atttacatgg tcacagtggc 122640
agggaatctt agcatgattg ccctcatcca ggccaatgcc cggctccaca cgcccatgta 122700
ctttttcctg agccacttat ccttcctgga tctgtgcttc tcttccaatg tgaccccaaa 122760
gatgctggag attttccttt cagagaagaa aagcatttcc tatcctgcct gtcttgttca 122820
gtgttacctt tatatcatct tggtacacgt tgagatctac atcctggctg tgatggcctt 122880
tgactagtac atggccatct gaaaccctct gctttatggc agcaaaatgt ccaaaagtgt 122940
gtgttccttc ctcatcacgg tgccttatgt gtatggagcg ctcactggcc tgatggagac 123000
catgtggacc tacaacctag ccttctgtgg ccccaacgaa attaatcact tctactgtgc 123060
agacccacca ctgattaagc tggcttgttc tgacacctac aacaaggagt tgtcaatgtt 123120
tgttgtggct ggctggaatc tttcgttttc tctcttcatc atatttattt cctacttta 123180
catttttcct gctatcttaa ggattcgctc tacagagggc aggcaaaaag ctttttctac 123240
ctgtggctcc catctgacag ctgttactat tttctatgca actctgttct tcatgtgtct 123300
cagacctcca tcagaagagt ccatggagca aggacaaatg gtagctgtac tttataccac 123360
tgtgatcccc atgttaatcc catgatctac agtctgagga acaaggatgt gaaaaaggct 123420
ttatccaaag aactgttcaa aagaaaattg tttcctaaat aaacatcagt attgattttt 123480
gtcatgctgt cattttattt agcctataat tttttcatag agcttagtgc aatacaaatt 123540
tatccaaaat tatacatttt ccttgaaggg ttagggaatt tttatgaagt gtgaataaaa 123600
gaaatatgag tatttacatc aattaacaga tagtttgata tcaatatagt tttaactgcc 123660
catacccaa agtaaaaatt tctatagtga ggaatcaatg taaataaaaa aaaatctaat 123720
ttgtatttta gaaaataaaa cccttaaaac cagacctggg ttttcttttt ggaaggagtt 123780
aggcatatag ttttaaactg cttcttaaac atatattaag ctatttttt tttttgaaac 123840
agcatctcac tccatcaccc aggatggact ggagtgcagt ggcatgatca cagcttacag 123900
caaccttgat ctcagggttt caaaccatcc tcccacttac acccctcacc cctgccacag 123960
tagctgggac tacaggtgtg caccaccata ccacactaat ttttgtattt tttgtagaga 124020
cggggtttca tcatgttgcc cagactggtc ttgaactgct gagtgcaagt gatctatctg 124080
cctgggcctc caaaagagtt tgcattacat gggtgagcca ctgtgcccag cctgtaataa 124140
ggtttaaaaa agagcacttc tgctaaactc tactagatac ttttgcctat cttacagagg 124200
aagtcagcat tccactttac catgtccaac ctaaaacgta agcctcatcc cttgcacact 124260
cagttgtcat ctcagctgtt tcctctgcct gtaccaacct catctccagc ctaaaacaca 124320

ccaatgcttt ggtccacagt gaggcacact tatgtttcca cagtccatat ctggtttatt 124380
tcttggcttt ttagaggaaa aataaatgac cagtaggatt ctaaattgaa tgttacctaa 124440
aagtctgtct aaggagaccc aaggctaaca gataatctgt tcaatgcatg tctaaaacca 124500
catcaaccca aaaggtataa catgtggttg taatgatctt gaagatgtgt gtttttgggt 124560
gtggaggcct atagttcccc tagaggtaac ctagaactta agatataatc acatacaagg 124620
aagacgagag actaatgatg aggtgtgtgc aaggtgagaa attagaacag agatgccctc 124680
aggaaattcg gattattgaa gggctgcaac ctctagagta ggacagagtc aacaacagca 124740
acaacaagga tgaaactagt ccgtttaggc tttgtttcat ggaggagtta aaaagtctta 124800
tttaagaatt tatgacaata gattaacggt catattaatt cagcgtttat atttacactg 124860
ctttgatggg ctatcaaatc tgaagctaaa aaaacatgaa ataggcccgg cgtggtggct 124920
cacacctgaa atcccagcac tttgggaggc ccaggcggtt ggatcacgag gtcaggagat 124980
tgagaccatc tgcctaacac ggtgaaaccc cgtctctact aagagtacaa aaaatcagca 125040
gggcgtggtg gcgggcgccg gtagtcttag ctactcagga gactgaggca gaagaatggc 125100
ctgaacttgg gaggcgcagc ttgcagtgag tggagattga gccactgcac tccagcctgg 125160
gtgacagaag gagactctgt ctcaaaacaa acataaacaa aaacaaaaac aagcaaacaa 125220
aaaacaaaaa caagcaaaca aaaaacaaaa cataaagagg cccaggttag aagaaacctc 125280
atgacactgg aataagcaaa tataaacata agtcagagaa tgaattatca actactcaga 125340
cttcaaagat atcccgcaga taagtttcta attgacagac attttccagg aaaataagaa 125400
ataaaaagca caaatgaata ccttcaactt attacagata ttaggattac ttgatgcaaa 125460
ttataaaata aagatattta atcattgaaa taagtgtgaa aattaaaatt cgaagaccat 125520
atatatgtgt gtatatatat gtatgtgtac atatatatgc aaagactgat tatataaagc 125580
aataattata tataatatgt agagttccga gcatatctgc aggccgaagg agagactctg 125640
tgttttatag gcattgaagc attttacatt ttttttttatg ttctcagcac ctttatttat 125700
ttatttattt ttttagtatt tattgatcat tcttgggtgt ttctcgaaga gggggatttg 125760
gcagggtcat aggacaatag tggagggaag gtcagcagat aaacatgtga acaagggtct 125820
ctggttttcc taggcagagg accctgcggc cttccgcagt gtttgtgtcc ctgggtactt 125880
gagattaggg agtggtgatg actcttaacg agcatgctgc cttcaagcat ctgtttaaca 125940
aagcacatct tgcaccgccc ttaatccatt taaccctgag tggacacagc acatgtttca 126000
gagagcacgg ggttgggggt aaggttatag attaacagca tcccaaggca gaagaatttt 126060
tcttagtaca gaacaaaatg gagtctccta tgtctacttc tttctacaca gacacagtaa 126120
caatctgatt tctctttctt ttccccacat ttccccctta tctatttgac aaaactgcca 126180
tcgtcatcat ggcccattct caatgagctg ttgggtacac ctcccagacg gggtggcggc 126240
cgggcagagg ggctcctcac ttcccagacg gggggccgg gcagaggcgc cccccacctc 126300
ccgggcgggg tggctgctgg gcgggggctg ccccccacct ccctcccgga tggggcggct 126360
ggccgggtag gggctgcccc ccacctccct cccggacggg gcagctggcc gggcgggggc 126420
tgccccccac ctccctcccg gacggggcag ctggccgggc aggggctgac ccccacctcc 126480
tggacagggc ggctgctggg tggagacacg cctcacttcc cggatggggc ggctgccagg 126540
cggaggggct cctcacttcc cagacgtggc ggctgccggg cggaggggct cctcacttct 126600
cagatgtggc ggcggccatg cggaggagct cctgacttct caggcagggc agccgggcag 126660
agacggtcct cacctcccag acggggtggc ggtcgggcag agacactcct cagttcccag 126720
acggggtcgc ggccaagcag aggcgctcct cacttcccag actgggtggc cgagcagagg 126780
ggctcctcac atcccagacg atgggcggcc aggcagagac gctcctcact tcccagacag 126840
ggtggcggct gggcagaggc tgcaatctgg gcactttggg aggccaaggc aggcagctgg 126900
gaggtggagg ttgtagcaag ccgagatgac gccactgcac tccagcctgg gtaacattga 126960
gcactgagtg agtgagactc cgtctgcaat cccggcacct cgggggggctg aggcgggcag 127020
atcactcgcg gtcaggagct ggagaccagc ccggccaaca cggcgaaacc caccaaaaaa 127080
tacaaaaacc agtcaggtga gcggggcatg cctgcaatcc ccggcactcg gcaggctgag 127140
gcaggagaat caggcaggga ggttgcagtg agccgagatg gcggcagtac agtccagcct 127200
cggctgggca tcagagggat accgtggaga gagagggaga gggagaggaa gagggagacc 127260
gtggagggag agggagaggg agaccgtgga aggagaggga gagggagagc attgaagcat 127320
tttagagtgg ataggtaggt atataaattt tcacattgaa aggaagagga agaaacagct 127380
aagtccctct catgagactg tcttggagag ggttacagca gagaggctgc ctttctctcc 127440
ttcctcctca gtaaattttc tacagacctt ggcagcctaa tctctacagt cctgcagcag 127500
tgagtcctct gaactggcac catagttcac tgggaaaggc attcatccat caagttcaga 127560
gaggtggcca tccatgttgc taaacagaat aaaacaaaaa ttcactttgt catctgcctg 127620
acattggcct tgctatgcaa tagcatcata aactgcctga aaggtcttat tattttcaaa 127680
ggattttgtg ctgtctacaa gtggtatttt tctccacctc agaaataaaa tattttagga 127740
ttgacccacc tcaaatttcg gtttgtcttg gttatcggcc actgcttata taagtctaat 127800
tgtgtgacag aagttcccaa tgctctgctt catgaagtac cctttaggac tacctcttgt 127860
ctcttcagtt cttcaatcaa ctttttcctg aagttctcgt gtggggatct gatataaaca 127920
gtaggcagct gctggcatcc agcccagagc cacctcttcc aggcccagga catccttaaa 127980
gcacctctga ccagcaacct caatctggcc aacacttttt tcagcatata caggtaaatc 128040
tccatgggct tgatcagcac cttcacaaat cagaaaagaa accacaaccc agagtaaaga 128100

aaaaacaacc tttcaaagtt agtaaggcta ttggtggtgt taaaatatca cccggttaaa 128160
aaacaaacaa actgatagca gaaaaaaatg gcaaaatatg aactggtgct atataaaaac 128220
atatttaaag aaccaagtaa cttatgactt tgtgctcaac ctgaataata attagggaag 128280
tgtacattac aagatatctg tatctttata tatctgtacc tatatatctg tgtctgtatg 128340
ttgcagaatg acagaaaaac tgatattgtt aagcattgga gagactgtgg aacaactgac 128400
cctctcatat aatgctagtg ggaatgcaaa tggatataaa cagtatggaa aactactagg 128460
cagcataaat ttaagctgca catatttata ctccataaac caacactact gcccaaagta 128520
ctaacaggca gaaatacatg aatatgttcg tgaagagata tgcacaagat tgaattttgt 128580
aacagccaca cctggaagta tcctatatgt gtatcaatag taggaaggat ttattaattt 128640
tggtgtactt actatacagc cgtggagatg aaggaattat ttacatattc atgacactcc 128700
tcattttgaa taaaaatcac caacataata tttagcacca atgtagtgtt cattggaagg 128760
ctagacctaa cagaatatgc actgtataat ttctattatg gaaaactcaa aagcagacac 128820
aattaatgta cagtgttaaa agtgaagatg tagttaattt gaggttagtg actaggaaag 128880
ggcaagcaga gagatagtgc agtactgtta atgttctact tgaggattta gaggcaagta 128940
atacaagtat gttcactttg taaaaattca tgaagctgta cgcctacatt ttgtgcacac 129000
tttctgtgtg taaattagac ttcaatataa aggttactaa aaacgaataa aaatagtact 129060
agacttcaag caagtaaagc ttcattccaa tatcaaagca ttctatttac ccatcagtac 129120
acagagggta ttagtttgct agggctgcca caaataagta ccatgaactt ggtgacttaa 129180
acatgcagat ttatttcctc acagttctag aggctagaag tccaagatca aggtgtgggc 129240
aaaactggtt tcattctgag ttctttttct atcttgtgga tgatcatctt atcccgacct 129300
ctttacactt tctttttctg tgtgtatctg tattctaatc tcttcttata aggatgcaag 129360
tcatattgga ttagggcaca gctcacccgc tagaccttat cttacttaaa tgttctcttt 129420
cgatatcgtg tctccaacag tcacgctctg tggggcttgg agctggaact tcagcataag 129480
aattttggag agtgagggaa gaggcacaat ccagtccata acacagatta agaaatgtaa 129540
aatgctaata gaattttgac acaaagtttg tgacactggt aggagagaaa ctgtatcaga 129600
aaagttgaat taagttgaaa gtaacatggt aaacctaagg caatgtgaga atccatggca 129660
gtcaggaatg ttattgtatg gattttctaa tgtaagaagg aaaatgctga gactgaaaca 129720
taaggcagag aaggaccaga gagttgtgag ttcccatttt aaatttgtgt tgtgccaaat 129780
gtcatctctc tagagaaatt attcagtgag aaaaaaaatc taacagagta attgcttcat 129840
ttttgcatat ctgtgaaatc ccttagggaa ataaagtcat catacaaata ttataaatta 129900
ttcctgtatt tgtcaccaga aaagccattg atattctttg taaggacagc tcttccctta 129960
ttcataggta agtttctgca tgtgttttta atcctggaac tctacttgct atacaatcgt 130020
atgtattttc agagttagat atatgattgt gatgattaaa tgactaggta gaaagaaaaa 130080
tgccaattac cagaaaaatg tagacagtta gtatttaaga tacttttatt tgttaaagtt 130140
ttgattaatg aggatggaag ttaatggcat aaaaatataa gaggcatgct ctaggatctt 130200
tcactcaata taaatgaaag ctaatattta ttaagggttt accacacatt gggcacagtg 130260
ctacgcatat tacatactcc attttgtgaa atcctaaaaa tagcactttt gtgtttgtta 130320
atttcatcag taatagaaaa aaactatagc ccagagttat taagaaatat gacccagact 130380
actcagatca gaagttctga catcacagtg tgatctcaac cagttgactc caaagcacat 130440
gtttctacca gtacagtatg ctttatggtt cgtagtggaa tttccttctg tactaaccat 130500
gagggaaata tgctattatc catacctatt acaggcagag tgtcatagaa tcggtttgag 130560
ggttgaatgt gttaaaactt ataaaataga ttagcgtttg gattataaga aacaccatgt 130620
aagtgctggt taaattagtt gtaaaactaa aacacagaat aaggaacatg tcaaaagaat 130680
agagcagcat ttcagaaata tctaactcca gatcctgtga actgatttta tgctaagcct 130740
attgcatttt tatcaaagca ttcatctttt tgtttggttg agtcctaaaa cttaagaatg 130800
tcaaccagat gtgtgcattt gtcaaacaca caaagttggg cactttaaat atgtgaattt 130860
cactgtatat aaattattgc aaaaacaaca ttaaacaaag aaacaaaggt gaaatctgac 130920
aggagcttga tatataaaaa tgaatgaggg tacatcacta ctatgattga atagatgtag 130980
acacagcttt tactcaatgt tgtataaatc acaaataaaa tctttgtttc agatattcaa 131040
aaataacctt ctataagttg ttcttgtgaa tatagatgat ttttatatag aaaaatagaa 131100
ggtacatttc attaatatca gtgttaacag tttttgttgt tgttgtaagt aacagaacac 131160
agctctggtt atgctaagca aaaccggaac attctgaaag gatgttaggg ctcctaaatc 131220
aatgtgcggt taaaagacca ggctcagaga acagatagca gcctaggcag gtgtggaggc 131280
tgagcagatc aaaccacaca gaatcagtgt ttctttgcaa catcagcctg atcttcaaac 131340
ttcactgtaa tcaattagat gaaaaaatat tttgcaacct tgagtcatga acttacttaa 131400
gacaagcagg gactaacatc tgacccctgc ctactccttt atggtttttt ttttttttt 131460
ttttttttt ttttttttt ttttttttg agacagcatc tggctctgtc gtccaggctg 131520
gagtgcagtg gcaccgtctt ggctcactgc aagctccgcc tcccgggttc acgccattct 131580
cctgcctcag cctcctgagt agctgggact acaggcgcct gccaccacac ctgccctgac 131640
tactcctttc taaacagaga atctggtctc tcagtggaaa ctagaagtgg aagatcaaga 131700
ggagagaatt tccatgagtc ttaactccac ataatggaaa atttcctcca aaaaattcaa 131760
agagagtgtt actggctaga gaatttaata gctgtgcaac cctcaaaagg caaatatgca 131820
gacacttttc aaaataaggg gaaaaaaatc atgccctgcc tggagaatcc ctattgcact 131880

gagcctggag gaataagaaa gcaacaaatg agcccaacct gtttgctcag gcttctccag 131940
gaagggatcg gattggtgaa aactattcac aaagaacaag atcctcgggc aaatattgct 132000
ttttgatact tgtccttttc tctctttacg cgcattctct gcccagtatt ataatcggga 132060
aaaaagaaat tgcttagttc tgaatgttat gttttggggc aagtccacct ataaagtaca 132120
tttctcaaat aaaaggagag gcagaataag gggaggagaa acgagagaag aagaatagaa 132180
aaacagaagc tgaatttgaa gggagtgaaa agaaagaaaa tggaaagaaa acaagaaaaa 132240
taaagggcag tatcataaaa ggtagagagc tgctttgaat atccctaatt tagtttacct 132300
gaaacaaagt tcagtagaaa tagaatgagg aaaactgcca aaaagagggg acgtttttgg 132360
taaaagatca ttactgtaat agtaataatc atgttgctat tgtattgtaa tccaatatat 132420
atgtaaagtt cctttcttcc accgaaggaa attatgagaa gaaactgcac gttggtgact 132480
gagttcattc tcctgggact gaccagtcgc cgggaattac aaattctcct cttcacgctg 132540
tttctggcca tttacatggt cacggtggca gggaaccttg gcatgattgt cctcatccag 132600
gccaacgcct ggctccacat gcccatgtac tttttcctga gccacttatc cttcgtggat 132660
ctgtgcttct cttccaatgt gactccaaag atgctggaga ttttcctttc agagaagaaa 132720
agcatttcct atcctgcctg tcttgtgcag tgttaccttt ttatcgcctt ggtccatgtt 132780
gagatctaca tcctggctgt gatggccttt gaccggtaca tggccatctg caaccctctg 132840
ctttatggca gcagaatgtc caagagtgtg tgctccttcc tcatcacggt gccttatgtg 132900
tatggagcgc tcactggcct gatggagacc atgtggacct acaacctagc cttctgtggc 132960
cccaatgaaa ttaatcactt ctactgtgcg gacccaccac tgattaagct ggcttgttct 133020
gacacctaca acaaggagtt gtcaatgttt attgtggctg gctggaacct ttctttttct 133080
ctcttcatca tatgtatttc ctacctttac attttccctg ctattttaaa gattcgctct 133140
acagagggca ggcaaaaagc tttttctacc tgtggctccc atctgacagc tgtcactata 133200
ttctatgcaa ccctttttctt catgtatctc agacccccct caaaggaatc tgttgaacag 133260
ggtaaaatgg tagctgtatt ttataccaca gtaatcccta tgctgaacct tataatttat 133320
agccttagaa ataaaaatgt aaaagaagca ttaatcaaag agctgtcaat gaagatatac 133380
ttttcttaaa aatcagtatt cttttggttt ctaaagccct tcctagactt ttttctttag 133440
ctgagaaata tagtgcatca atggagaaca ttgcagtttt caaaacttta tttatttta 133500
ttttattatt atattttgag atggagtttc tctctgtctt aggctggagt gcagtggtgt 133560
gatctcggct cactgcaacc tttgcctccc gggttcaagc aattctcctg cctcagcctc 133620
ccgagtagct gggaacacaa gcgcacacca ccatgcccga ctcatttttt gtattttagt 133680
agagacaggg tttcaccatg ctggccaggc tcggcctccc aaagtgctgg gattacacgc 133740
atgaaacacc gcgcccagtc taaaaacttt attttctaaa attcaaatac gtacaatttt 133800
gttcacaaaa gcttttatgt tttaagttgt cattcatctt gttcagcagt tattttaagt 133860
tctttgtctt ccgtgcacag aatggctttg tacctccatg cccttaggtt taggtaagat 133920
catgtggcta tttctggaag atgagattaa aattcacata tgtcacttct cagttgaata 133980
cttaattgat tgttagarca tttttgagtg gatctgttgt ttttatttt tccatggtga 134040
ccagcaagac ttaatatagt ggctgcccag tcaggagaca tggaacactg ttctcatccg 134100
actcacaatg gcaatacaaa acggctgata aataacgttt tcataagcta ctatgatatg 134160
tggagtttgc ttgtttccct agcacaaaat agcctattct gattgatgcc atattaatta 134220
tacaaattca cacaggctga attatataat atagaatgag aacatcacaa accttgaaga 134280
aatcacattt gagattttgt tttgtgttct tgtagtcatt tttaatgctt ttacatacac 134340
atgtgtgttc agttaaaatt atttattata attaccattt gaacctactc atgattaatg 134400
tatgtagttt taaatgtcaa ataatattac aagccatata taaactcctg tctcacttct 134460
accttaccac taactcccca tagataatac tttcttgtta tttcctttgg aaattaaccc 134520
attttaaatt atatgcttgt attcctctgt actgattttt caaatatgca cataatttat 134580
tgacatctgt ttattaaaga taaggtgcca tcattctttc accaatcact tcattcttac 134640
cctacaatga tcacttcttt ctcttcattg tcacagtagg atcacatttg aatttttggc 134700
ttaatcagta tctactgtta aaattactat gaatgtatca ttaataatac ctgaagtata 134760
ctataatact ttgagcatca ctaacactag tattaattat taataatatg aatactacta 134820
tgattgtaat tctcttcttg cacagttact ggattttcat aaagttgaca atggcctcca 134880
gttttgctct tctttttgta tagtagctac ttttccccaa actgtcacta aactgtaaag 134940
caccttttata gggtcaaaca caatagatga tctatagttt ccatattttt ttctcagacg 135000
ttctctgtcc tcccattgga ttctgctctg tttgcagagc tgttatttga agataatgat 135060
tccaatcatt ggtttaaggc atttcattca ttcaacttga ttacatgaca atatgcaaca 135120
aactggagat tcaacaaaaa taagacagaa aaaagtctta aagagagttt tatggaacag 135180
ttaatgaact tttctgctgt tatcatgagt ggcacaaggt cagagataac taggtccatg 135240
catgtttgta tttttctgca atgtcagaat tttcttgatg ctatttcaat cataaaatcc 135300
atgagctaca tggggttccc gaggaggcaa ttctccttaa aacttcctgt tcactcagta 135360
gtcagagcca tgggcaaaca ggctcaagtc attcaacaag tcagtcagta ttgcagtcca 135420
tgaattacag tatatttaat cagtctataa atgtaataga ttacacattg tacagcaaac 135480
aaagtaacat ttatcatcaa gaataaaggg actggaaatg ggttaaggaa ccagtccact 135540
gagagtgaca tgaacaaaaa gaatagccta ctcaggccct ggtggtccat caataaccttt 135600
gtaaagaaga gttttcgttg tgggcaaagc cttcagtggc aaatgcaaga tgcttatctc 135660

aagtgataac aagatggtgt cttttaaggt ggctgtttca agctgctgaa atcctgctct 135720
tttatggaca cagagtcctc tagtaagaac tgatagtgga agagtgactt tgttatgtcc 135780
ttatctggtt ggatgcagtc tttcttgatt aggcaaaaca tctggccctt gttggcatga 135840
tgcctttaaa aatgtaagat gaagtcattt tctaagatgg agtaacttat atcaatgggg 135900
ctctatacta cgattcagcc ccaggttccc ttctacactt tcctttaccc ttaccatttc 135960
agccttagct gaagagctgg gtgcagtggc tcatgtttgt aatcccagca ctttgggagg 136020
ctgaggtggg tgaatcacca gaggtcagga gttcgccact agcctggcca acatggtaaa 136080
accctgtttc tactaaaaat acaaaaatta gctgagcatg ctggtggaca cctgtaatcc 136140
cagctatttg ggaggctgag gcaggagcat tgcttggacc caggagacag aggttgcagt 136200
gagccgagac cgcactgttg ccctccagcc tgggcaacaa gagtgaaatt ctatctcaag 136260
aaaatagact tagctgaaga attatctctt cctgcaaggt ttctttaaca cattcaggat 136320
gtgtagtgtc aactgaggaa tgatgagatt cataaattta gaaaggtgga ctttctcata 136380
aagggttgta gcctgtaggg tgaccgttct gacaggctgt gaagcatatc ctccagctag 136440
aagttggaaa gagacacttc gacagtatga agagtaagac agggatttat gctgaatggg 136500
atgaccaaat atattacaca tatatgataa tacatattca acaggctata gaaaaaacta 136560
tgaatattca caaagaagag gcacaggcat gaatagtagg ctaatataag caacatgcat 136620
cccatgttcc ctttggagtg gggacttaac atttaaatgt gtcatgatta gactctatgc 136680
accaaaaggt gaatcagaag acaccaagac cctctgtgca cagcctctgt tgactggcaa 136740
gagccactag gttattggtg gtctcttatc aagaaggaat gctggtcaat tgctgtgttg 136800
aaaccgcaaa aagagaagtc cagtgtcagg tggtttgcag atatcagtgg tggtgcgagt 136860
ctcacaaggg caggtttctg tttaaccctt attgtaggaa gcctaatggt gtttagcaag 136920
ggaggggaga taacgaggca tgtctgatct cccatctgtc atggcaggaa ctcagatttt 136980
aaagttttc tgaggtttcc ttgaccaaga ggcagtctgt tcaattgatt caggggttag 137040
gattttattt gtatttctca ttaggtacca catgatgatt cacccacaga ttcacaatta 137100
tattattgca ttcataacat ggttctataa ttatatgcat gtaaatctgt tccttccact 137160
attttagcaa gttctcaaag gaaaggacca catctttttg tttttatatt tttaccacct 137220
taagatagtg ttctataaag ggaggatgcc catttttttt ttgaaactgt gagaacaatc 137280
ccttccactt tctacctttg tctgtaatta tggccaatta cagatttctc tccatgatac 137340
tcgcttctcc catcctaaca tatattcaag gcagaacaat agatcattta gtttaagaaa 137400
accatgttca agttctttat cataaatggc ccactgaaag cccagcaacg tgaatcataa 137460
ctgagcaaga attggagaaa gtaatttcat tggcagcaga caggaaagat cacatactac 137520
atcctattct tcatagcaga gagacagata acaataaatg ctgaactaca gtaaaagatg 137580
ttaagggaaa tatttgtgag gaataaatct ttgtagcaat gtatttttcct tgatatgcaa 137640
tcataattat ccatagcatt tgggaaacaa ctgacaattt ttatcacctt tataaatgtt 137700
agttttgatc tttatagcaa ggttataatg gaaaaatcaa ccactgtgta ataaaattat 137760
tttaaaatga acagaattac actaggctgt ctgggacaga ggcaagggaa gggctgagtc 137820
atgatttaag gtgcaggaaa caaaagggac ctcattgttg ctcagacaga aaagaggttt 137880
ggtgggaatc agacaacagg tatatattga gacaacgaaa tatccaatcc ttgaaaaagt 137940
acattcttgt gcatcacttt ttcatagcca accgtcctaa gatttatgcc atgtgataag 138000
ctgatgataa aacattctct tcaagttgaa acagaataca gttgtggaaa aatatttggc 138060
tttgtaatta ttcagaactg agcctgaatc ctagttttat tactattttg ttagctgggt 138120
aaacttagag aagttccctc tccatatcag tttattcaaa tgcaaaaccc catttcatgg 138180
agttattgtg aatatcaaat attatttta atatacattt tcctcaattg cattttgagg 138240
caagtatgct gagttccagt gtggactcaa tttcatgaaa gtttttttaac atgggaaaca 138300
tgatcaaaac aataagttaa atatgttagt tattcattta tttaacatat gattattgtc 138360
cctgtcagtc actttcatca ttggcagtca ccagtctctt ctacctgttt cattgtttct 138420
tcatcagtct cttcctgttg tattgtttca tctatctgtt tctaaaacat ttcatgtttt 138480
ttccagaaat tatttgtacc tagtattatt tgtttacata tcttaggcat cccatcaaaa 138540
tgcaaatccc tttgttgaag aaatccttta aaaatatttt tattattcag attaaatagt 138600
attgaagttt gtacttagta catattcttg tgtgaacttg ataaaggaca aacaatggag 138660
gaaatatggt agtgctcttg agatgggaag gacaattacc tgaattcagg tttcatgaga 138720
agttaagtgt cctcagacta ttttaaattt gtttttcagt attgtataaa gtgctaccct 138780
cgttggataa caggtgattg ctttgagctc tactgaagta taaaatatag attttttttct 138840
aacatcttga ttcaagataa aatagtacaa ccaattagta tttcccttga gaatgttctc 138900
cactgtgatt tcacttctaa ttacctcttc ttacacattt ggggattctc ctgagatata 138960
tgctgatgcc agtgatgtgg gtatatatct tccagcactg gccagaattt gacctttggt 139020
gtcatgaaaa agacaccctt attttcacag gtaagaaagg aaatgttttt gcatattatt 139080
tttcagtatc taataatgtc tgtattcact ttgttaaatt gtgtacattt tagacagcaa 139140
tgtgtttaaa taggtagttt tgaggcaaat ttgcacagac aatgcatttt ccatgaaaga 139200
aaatctgtag ggagcttact tagcttcatc tttcactttt ttgatattac cagtgttatt 139260
gaaacatcct ggtggaattt catagtatgg ctatttcagc agatgttttc actgttatta 139320
gtttcaagta atattatttt tgtgttttcc atttgcctaa tgtcttggtg cctactgtca 139380
ccaatcgcca caagaaaaag gcaagttgaa aataatttaa gtctacatag cttaattaat 139440

cacataaaat ttttctgacg tctaagtaaa ataaatgctc aaacaggctt tttattgtga 139500
ccattaaaca tgtgtgaaag tggtccatta gcttcagtat gtggagctaa cgaacttggg 139560
tgagtagatc ttaagttctc ttgggttctc ttattttata agattctggg ttcttttcca 139620
atgccctatt tttgtctaga gacctagttt ttttacaagc aatacttgtg cctcagtgag 139680
actttagttt tcagaatatt attttacttt tttcaacaaa atacttgaaa cataaatcag 139740
cctgattata atcaaatcag tctgattata atcaaatcag tctgattcta atcaaataac 139800
taacttgtgt caaggacctt aaacagtaag ttgatgaatg aaattctcat acctatttt 139860
ctaggctgca caagtagtca cagaatcatt ctcgggtcat tatcaaattc atcatcatca 139920
aagttattga ctgcatttga atgtttgttt tgagtgtttc ttgcctctca gcagcagtgg 139980
acagatttct catgggctct atatctgtat attagctgtc tctgattgcc attgtcccgc 140040
aatgctagaa ttaaaatatg gttaaygata atggtctcca ctaaaagtgc atgtccagat 140100
gaccagtgaa acctgagcaa atttgcaaag ctccttagag actaagtcag gactaaaacc 140160
aagaccctgg attgaagaag ttgctcacca ccagagagtt aaaagaagcc cttgtgcctt 140220
ccttcttgcc aattaagtta caattcatag gcattattct agattcttgg cccagatttt 140280
tcatattctt tacagtttcg gcactgatac ctcggctggc tcgctttctt cttcctcatc 140340
catctcttat ctttgttaga gtgcttgtca ctatcatctg taactcaaag acaactctaa 140400
ctcaaagtcc tcatttattc atatttgttt cctgggacag gataaggagg caataatatt 140460
agataaattg aattttaaca ttttggaaac aatttctgtc attacccaac cagtcaatgg 140520
aaccaaagga agagatagac ctttctggga aagtaaaata ttacagaagg agagcaccgt 140580
tctgagaatt ataatatatg atttgtcttc tgattcttct gcttatgagt tgtaagctgg 140640
gaccagaatg tcaaattttc tgagcctggt tactcaccta actaaattta tttctactat 140700
attgtccact gctctatttc tagcatccag caccatgtct ggaacttgaa aaatgctcaa 140760
taattatctg ctataagcac gcatgcatac acaaatgcat ggctgacaac atattacaaa 140820
gtaaattagt gtttttttaa ctataaaatt tcactttaat aagagtttag caattgttgt 140880
aagatctcat acaatttaat agaaaatatg tgtctgatat tcttagatta ggtgaaaaat 140940
aaagtgctac tcgtattgac ttcctgagac tgctcagaag agataagttt tatcagcttc 141000
tgacaagttc tccactaatg acatttgctc atgtctgaat tgccccatgg aaaaaaatta 141060
gatgtaaatt tcattttacg caattgttag tatactgaaa ttgtttttaa ccaaaatttg 141120
gagacagatt aacacatatg atattttaac ttaatttgat attttatata tttaarataa 141180
tttgtgataa attcagtaaa tgcataatgt ctatataaat tgtaagctac aaaatataat 141240
gaatacacat taaaccactc cctagaatta taagtaattt ctgtgcattg tttaaagtaa 141300
atgcacaatt taataatttt cccatttaaa taatctataa tccattaagc caatagtaaa 141360
tactaaaata gtccctcttt ccaaccctga tctgaaatgt catctttgtt atatgccaca 141420
tcacctatat acatgtagct gttcctggac tttctattct ggaacattag ttaaagtgta 141480
tatttttttc caatttcatg tcaataacac actgtcctga atactggttt tacagcaaat 141540
cttgatatct aataagtaac tctgcttatt gccctgagag atattgagtc ttacggttat 141600
tttatacttc cgtataaata ttagaatcag ctattccaca agaattgtgt ttgaattttg 141660
attggaattt atttgactct atgggtcact taggaaagac tgacctttt atgctattga 141720
atccttccct tagtgaacaa agaatctctc tccattttta gttctttctt tctctctctt 141780
ttctttcctt tcctttcctt tcctttcttt cctttccttt cctttctttc tttcttcttt 141840
ctttttcttt ctttctttct ttcttccttc cttccttcct tccttccttc cttccttcct 141900
ttctttgttt ctgagacgaa ctctcgcctc tctttgccac ccaggctgga gtgcagtggc 141960
atgatctcag tgcactacaa cctctgcctc ctgtgttcaa gtgattctcc tgcctcagtc 142020
tcctgagtag ctggggttac aggtgcacac ctccatacct ggctgatttg ttttatttt 142080
agtagagaca gggtttcacc atgttggcca ggctggtctc gaactcatga ccagaagtga 142140
tcaacccgtc ttggcctccc aaaatgctgg gattataggc ataagccacc acatccagcc 142200
cattttaag gctttttaat gttgtgtggc aatggcatca aagtgatcac ataatttatt 142260
ctgggtatat tagcagggag taagtataga ttgagaacct gaaaacattt ttttaatcat 142320
cttagggtga ttgagttaat tccttttatc ctaatggcta aactacacat cagggttaac 142380
ataatggagg aataggaaat tctttgatat aggacttctt tatacagaga tataaaagag 142440
catgtgagcc aagcagggga ctataatgtt ggttctatgc agcatgaatg ttgtaaaata 142500
gcaagaaatt aaaatatagg aacaaattaa aacatagatt gatacctcaa agataataca 142560
tatgaaattt catatagaat tgagatattg atgaaggagc ataaatacac ttgtttaacc 142620
caacagacag acaggctaaa ttgggatgtg caaattggaa aaaatgctta taaacacgtg 142680
atatcaatgt ttaagactca aacgaaaggc tttatttaca taagttccca aaaactaact 142740
tataaacaaa gaaccaatat acatatcaca atgaaaagca ctactactaa aaacaacaaa 142800
atataaaaat acaccgagac agtaagttga actcggtgag atatatagta tattagcaga 142860
tcaccgaatg gggatgtatg agataagcat ctcagtggga ttagtcactt catctctcta 142920
ggattcagtc tctcagtctt cagaatgaag agctttgcct gcattatcct caaagtcctt 142980
tctacctctc catgtagctt tgtatgtata tgtaagcatg tcattagaag ttctaagcat 143040
ttcaaaatct ctattacaat ttaatcttta acatataagt tatctagatt tttaaaagaa 143100
atttcccaca gcataaggat actttcgttt ataattttgc attaagcctc caacttaatt 143160
gcacactgtg attacaaatt cagcctcagt ggaaaccact tgttgcaaat gtttagactt 143220

```
gcttcatagt ataactgttc atcttcagtt acagaactgc tactgagata acataactaa 143280
agccttttgg ctctttttat acaaagcatg atatttaact agggttttag tgatttttaa 143340
aaagtttctc tttctcctta gatattcaga ccaatgcgtc tcatatgaga tgaagaaatg 143400
tccagaagaa actatactga actgacagaa tttgttctct tgggtctaac aagccgtcca 143460
gagctgcgag ttgctttctt ggcactgttc ctttttgtct acatagccac tgtggtagga 143520
aacttgggga tgattatttt aatcaaagtt gattctcgac ttcacactcc catgtaattt 143580
tttctctcca gtttgtccat tctagatctg tgtttctcca caaatttcac tcccaaaatg 143640
ctagaaaatt tcttatcaga gaagaagacc atttcctatg caggttgttt gatgcagtgc 143700
tatgttgtca ttgctgtggt ccttgcagag cactgcatgt tggcagtcat ggcatatgac 143760
cgctatatgg ccatctgtaa tccattgctc tacagtagca aaatgtccca aggtgtttgt 143820
gtccacctgg tcattgtccc ttatgtctat ggctttcttc tcagtgtgat ggaaacctta 143880
aggacctaca acctctcctt ctgtggaaca aatgaaatca accatttcta ctgtgctgat 143940
cctcctctta tcaaactggc atgctctgac acgtacagca aggagctgtc catgtacata 144000
gtagccggct acagcaacgt ccagtctctt ctratcattc tcacatccta catgttcatc 144060
cttgtcgcta tcctcagaag ccattctgca gagggaagga aaaaagcttt ttccacatgt 144120
ggttcccacc tgacagttgt cacaatcttc tatggaaccc tcttctgcat gcatttgaga 144180
cgtcccacag acgagtccgt ggagcagggg aaaatggtgg ctgtgtttta caccacagtg 144240
atactcatgc tgaactccat gatctatggc ctcaggaaca aggatgtgaa agaggcgttg 144300
aaaaaagcaa taggaaaaca aacattggga aaataaaaat gctaagctat cattaaaaat 144360
ttgtgaagta atgagatata atatcattgg gttagatgtc acattttagg ctacatttgc 144420
acaattcatt tctaattttc tgttaggtag ctgactgagt                       144460
```

<210> 2
<211> 195
<212> DNA
<213> Homo sapiens

<400> 2

```
atg agc ttc tta ata aga agt gat tca aca cta cac act cca atg tgc       48
Met Ser Phe Leu Ile Arg Ser Asp Ser Thr Leu His Thr Pro Met Cys
1               5                  10                  15
ttg ttc ctc agt cat ctc tcc ttt gta gat ctc tat tat gcc acc aat       96
Leu Phe Leu Ser His Leu Ser Phe Val Asp Leu Tyr Tyr Ala Thr Asn
            20                  25                  30
gcc act cct ccg atg ctg gtt aac ttt ttt ttt cca aga gaa aaa ccg      144
Ala Thr Pro Pro Met Leu Val Asn Phe Phe Phe Pro Arg Glu Lys Pro
        35                  40                  45
ttt cct tta ttg gtt gct tta tcc aat ttc acc ttt tca ttg cac tgg      192
Phe Pro Leu Leu Val Ala Leu Ser Asn Phe Thr Phe Ser Leu His Trp
    50                  55                  60
tga                                                                 195
 *
65
```

<210> 3
<211> 948
<212> DNA
<213> Homo sapiens

<400> 3

```
atg ttc tcc cca aac cac acc ata gtg aca gaa ttc att ctc ttg gga       48
Met Phe Ser Pro Asn His Thr Ile Val Thr Glu Phe Ile Leu Leu Gly
1               5                  10                  15
ctg aca gac gac cca gtg cta gag aag atc ctg ttt ggg gta ttc ctt       96
Leu Thr Asp Asp Pro Val Leu Glu Lys Ile Leu Phe Gly Val Phe Leu
            20                  25                  30
gcg atc tac cta atc aca ctg gca ggc aac ctg tgc atg atc ctg ctg      144
Ala Ile Tyr Leu Ile Thr Leu Ala Gly Asn Leu Cys Met Ile Leu Leu
        35                  40                  45
atc agg acc aat tcc cac ctg caa aca ccc atg tat ttc ttc ctt ggc      192
Ile Arg Thr Asn Ser His Leu Gln Thr Pro Met Tyr Phe Phe Leu Gly
    50                  55                  60
cac ctc tcc ttt gta gac att tgc tat tct tcc aat gtt act cca aat      240
```

```
His Leu Ser Phe Val Asp Ile Cys Tyr Ser Ser Asn Val Thr Pro Asn
65                  70                  75                  80
atg ctg cac aat ttc ctc tca gaa cag aag acc atc tcc tac gct gga      288
Met Leu His Asn Phe Leu Ser Glu Gln Lys Thr Ile Ser Tyr Ala Gly
                85                  90                  95
tgc ttc aca cag tgt ctt ctc ttc atc gcc cta gtg atc act gag ttt      336
Cys Phe Thr Gln Cys Leu Leu Phe Ile Ala Leu Val Ile Thr Glu Phe
            100                 105                 110
tac ttc ctt gct tca atg gca ttg gat cgc tat gta gcc att tgc agc      384
Tyr Phe Leu Ala Ser Met Ala Leu Asp Arg Tyr Val Ala Ile Cys Ser
        115                 120                 125
cct tta cat tac agt tcc agg atg tcc aag aac att tgc atc tct ctg      432
Pro Leu His Tyr Ser Ser Arg Met Ser Lys Asn Ile Cys Ile Ser Leu
    130                 135                 140
gtc act gtg cct tac atg tat ggc ttc ctt aat ggg ctc tct cag aca      480
Val Thr Val Pro Tyr Met Tyr Gly Phe Leu Asn Gly Leu Ser Gln Thr
145                 150                 155                 160
ctg ctg acc ttt cac tta tcc ttc tgt ggc tcc ctt gaa atc aat cat      528
Leu Leu Thr Phe His Leu Ser Phe Cys Gly Ser Leu Glu Ile Asn His
                165                 170                 175
ttc tac tgc gct gat cct cct ctt atc atg ctg gcc tgc tct gac acc      576
Phe Tyr Cys Ala Asp Pro Pro Leu Ile Met Leu Ala Cys Ser Asp Thr
            180                 185                 190
cgt gtc aaa aag atg gca atg ttt gta gtt gca ggc ttt act ctc tca      624
Arg Val Lys Lys Met Ala Met Phe Val Val Ala Gly Phe Thr Leu Ser
        195                 200                 205
agc tct ctc ttc atc att ctt ctg tcc tat ctt ttc att ttt gca gcg      672
Ser Ser Leu Phe Ile Ile Leu Leu Ser Tyr Leu Phe Ile Phe Ala Ala
    210                 215                 220
atc ttc agg atc cgt tct gct gaa ggc agg cac aaa gcc ttt tct acg      720
Ile Phe Arg Ile Arg Ser Ala Glu Gly Arg His Lys Ala Phe Ser Thr
225                 230                 235                 240
tgt gct tcc cac ctg aca ata gtc act ttg ttt tat gga acc ctc ttc      768
Cys Ala Ser His Leu Thr Ile Val Thr Leu Phe Tyr Gly Thr Leu Phe
                245                 250                 255
tgc atg tac gta agg cct cca tca gag aag tct gta gag gag tcc aaa      816
Cys Met Tyr Val Arg Pro Pro Ser Glu Lys Ser Val Glu Glu Ser Lys
            260                 265                 270
ata act gca gtc ttt tat act ttt ttg acc cca atg ctg aac cca ttg      864
Ile Thr Ala Val Phe Tyr Thr Phe Leu Thr Pro Met Leu Asn Pro Leu
        275                 280                 285
atc tat agc cta cgg aac aca gat gta atc ctt gcc atg caa caa atg      912
Ile Tyr Ser Leu Arg Asn Thr Asp Val Ile Leu Ala Met Gln Gln Met
    290                 295                 300
att agg gga aaa tcc ttt cat aaa att gca gtt tag                      948
Ile Arg Gly Lys Ser Phe His Lys Ile Ala Val  *
305                 310                 315

<210> 4
<211> 519
<212> DNA
<213> Homo sapiens

<400> 4
atg tta aag aaa aac cat aca gcc gtg act gag ttt gtt ctc ctg gga       48
Met Leu Lys Lys Asn His Thr Ala Val Thr Glu Phe Val Leu Leu Gly
1               5                   10                  15
ctg aca gat cgg gct gag ctg cag tcc ctt ctt ttt gtg gta ttt cta       96
Leu Thr Asp Arg Ala Glu Leu Gln Ser Leu Leu Phe Val Val Phe Leu
            20                  25                  30
gtc atc tac ctt atc aca gta atc ggc aat gtg agc atg atc ttg tta      144
Val Ile Tyr Leu Ile Thr Val Ile Gly Asn Val Ser Met Ile Leu Leu
        35                  40                  45
```

```
atc aga agt gac tcg aca cta cac act cca atg tac ttc ttc ctc agt      192
Ile Arg Ser Asp Ser Thr Leu His Thr Pro Met Tyr Phe Phe Leu Ser
    50                  55                  60
cac ctc tcc ttt gta gat ctc tgt tat acc acc aat gtt act cct cag      240
His Leu Ser Phe Val Asp Leu Cys Tyr Thr Thr Asn Val Thr Pro Gln
65                  70                  75                  80
atg ctg gtt aac ttt tta tcc aag aga aaa acc att tcc ttc atc ggc      288
Met Leu Val Asn Phe Leu Ser Lys Arg Lys Thr Ile Ser Phe Ile Gly
                85                  90                  95
tgc ttt atc caa ttt cac ttt ttc att gca ctg gtg att aca gat tat      336
Cys Phe Ile Gln Phe His Phe Phe Ile Ala Leu Val Ile Thr Asp Tyr
            100                 105                 110
tat atg ctc aca gtg atg gct tat gac cgc tac atg gcc atc tgc aag      384
Tyr Met Leu Thr Val Met Ala Tyr Asp Arg Tyr Met Ala Ile Cys Lys
        115                 120                 125
ccc ttg tta tat gga agc aaa atg acc agg tgt gtc tgc ctc tgt ctg      432
Pro Leu Leu Tyr Gly Ser Lys Met Thr Arg Cys Val Cys Leu Cys Leu
    130                 135                 140
gct gct gct ccc tat att tat ggc ttt gca aat ggt cta agc aca gac      480
Ala Ala Ala Pro Tyr Ile Tyr Gly Phe Ala Asn Gly Leu Ser Thr Asp
145                 150                 155                 160
cac cct gat gct tcg tct gtc ctt ctg tgg acc caa tga                  519
His Pro Asp Ala Ser Ser Val Leu Leu Trp Thr Gln  *
                165                 170

<210> 5
<211> 948
<212> DNA
<213> Homo sapiens

<400> 5
atg ttg tcc cca aac cac acc ata gtg aca gaa ttc att ctc tta gga       48
Met Leu Ser Pro Asn His Thr Ile Val Thr Glu Phe Ile Leu Leu Gly
1               5                   10                  15
ctg aca gac gac cca gtg cta gag aag atc ctg ttt ggg gtg ttc ctg       96
Leu Thr Asp Asp Pro Val Leu Glu Lys Ile Leu Phe Gly Val Phe Leu
            20                  25                  30
gcg atc tac cta atc aca ctg gca ggc aac ctg tgc atg atc ctg ctg      144
Ala Ile Tyr Leu Ile Thr Leu Ala Gly Asn Leu Cys Met Ile Leu Leu
        35                  40                  45
atc agg acc aat tcc caa ctg caa aca ccc atg tat ttc ttc ctt ggt      192
Ile Arg Thr Asn Ser Gln Leu Gln Thr Pro Met Tyr Phe Phe Leu Gly
    50                  55                  60
cac ctc tcc ttt tta gac att tgc tat tct tcc aat gtt act cca aat      240
His Leu Ser Phe Leu Asp Ile Cys Tyr Ser Ser Asn Val Thr Pro Asn
65                  70                  75                  80
atg ctg cac aat ttc ctc tca gaa cag aag acc atc tcc tac gct gga      288
Met Leu His Asn Phe Leu Ser Glu Gln Lys Thr Ile Ser Tyr Ala Gly
                85                  90                  95
tgc ttc aca cag tgt ctt ctc ttc atc gcc cta gtg atc act gag ttt      336
Cys Phe Thr Gln Cys Leu Leu Phe Ile Ala Leu Val Ile Thr Glu Phe
            100                 105                 110
tac ttc ctt gct tca atg gca ttg gat cgc tat gta gcc att tgc agc      384
Tyr Phe Leu Ala Ser Met Ala Leu Asp Arg Tyr Val Ala Ile Cys Ser
        115                 120                 125
cct tta cat tac agt tcc agg atg tcc aag aac att tgc atc tct ctg      432
Pro Leu His Tyr Ser Ser Arg Met Ser Lys Asn Ile Cys Ile Ser Leu
    130                 135                 140
gtc act gtg cct tac atg tat ggc ttc ctt aat ggg ctc tct cag aca      480
Val Thr Val Pro Tyr Met Tyr Gly Phe Leu Asn Gly Leu Ser Gln Thr
145                 150                 155                 160
ctg ctg acc ttt cac tta tcc ttc tgt ggc tcc ctt gaa atc aat cat      528
Leu Leu Thr Phe His Leu Ser Phe Cys Gly Ser Leu Glu Ile Asn His
```

```
                165                 170                 175
ttc tac tgc gct gat cct cct ctt atc atg ctg gcc tgc tct gac acc      576
Phe Tyr Cys Ala Asp Pro Pro Leu Ile Met Leu Ala Cys Ser Asp Thr
            180                 185                 190
cgt gtc aaa aag atg gca atg ttt gta gtt gca ggc ttt act ctc tca      624
Arg Val Lys Lys Met Ala Met Phe Val Val Ala Gly Phe Thr Leu Ser
        195                 200                 205
agc tct ctc ttc atc att ctt ctg tcc tat ctt ttc att ttt gca gcg      672
Ser Ser Leu Phe Ile Ile Leu Leu Ser Tyr Leu Phe Ile Phe Ala Ala
    210                 215                 220
atc ttc agg atc cgt tct gct gaa ggc agg cac aaa gcc ttt tct acg      720
Ile Phe Arg Ile Arg Ser Ala Glu Gly Arg His Lys Ala Phe Ser Thr
225                 230                 235                 240
tgt gct tcc cac ctg aca ata gtc act ttg ttt tat gga acc ctc ttc      768
Cys Ala Ser His Leu Thr Ile Val Thr Leu Phe Tyr Gly Thr Leu Phe
                245                 250                 255
tgc atg tac gta agg cct cca tca gag aag tct gta gag gag tcc aaa      816
Cys Met Tyr Val Arg Pro Pro Ser Glu Lys Ser Val Glu Glu Ser Lys
            260                 265                 270
ata att gca gtc ttt tat act ttt ttg agc cca atg ctg aac cca ttg      864
Ile Ile Ala Val Phe Tyr Thr Phe Leu Ser Pro Met Leu Asn Pro Leu
        275                 280                 285
atc tat agc cta cgg aac aga gat gta atc ctt gcc ata caa caa atg      912
Ile Tyr Ser Leu Arg Asn Arg Asp Val Ile Leu Ala Ile Gln Gln Met
    290                 295                 300
att agg gga aaa tcc ttt tgt aaa att gca gtt tag                      948
Ile Arg Gly Lys Ser Phe Cys Lys Ile Ala Val  *
305                 310                 315
```

<210> 6
<211> 918
<212> DNA
<213> Homo sapiens

<400> 6

```
atg tcc aac aca aat ggc agt gca atc aca gaa ttc att tta ctt ggg       48
Met Ser Asn Thr Asn Gly Ser Ala Ile Thr Glu Phe Ile Leu Leu Gly
1               5                   10                  15
ctc aca gat tgc ccg gaa ctc cag tct ctg ctt ttt gtg ctg ttt ctg       96
Leu Thr Asp Cys Pro Glu Leu Gln Ser Leu Leu Phe Val Leu Phe Leu
            20                  25                  30
gtt gtt tac ctc gtc acc ctg cta ggc aac ctg ggc atg ata atg tta      144
Val Val Tyr Leu Val Thr Leu Leu Gly Asn Leu Gly Met Ile Met Leu
        35                  40                  45
atg aga ctg gac tct cgc ctt cac acg ccc atg tac ttc ttc ctc act      192
Met Arg Leu Asp Ser Arg Leu His Thr Pro Met Tyr Phe Phe Leu Thr
    50                  55                  60
aac tta gcc ttt gtg gat ttg tgc tat aca tca aat gca acc ccg cag      240
Asn Leu Ala Phe Val Asp Leu Cys Tyr Thr Ser Asn Ala Thr Pro Gln
65                  70                  75                  80
atg tcg act aat atc gta tct gag aag acc att tcc ttt gct ggt tgc      288
Met Ser Thr Asn Ile Val Ser Glu Lys Thr Ile Ser Phe Ala Gly Cys
                85                  90                  95
ttt aca cag tgc tac att ttc att gcc ctt cta ctc act gag ttt tac      336
Phe Thr Gln Cys Tyr Ile Phe Ile Ala Leu Leu Leu Thr Glu Phe Tyr
            100                 105                 110
atg ctg gca gca atg gcc tat gac cgc tat gtg gcc ata tat gac cct      384
Met Leu Ala Ala Met Ala Tyr Asp Arg Tyr Val Ala Ile Tyr Asp Pro
        115                 120                 125
ctg cgc tac agt gtg aaa acg tcc agg aga gtt tgc atc tgc ttg gcc      432
Leu Arg Tyr Ser Val Lys Thr Ser Arg Arg Val Cys Ile Cys Leu Ala
    130                 135                 140
aca ttt ccc tat gtc tat ggc ttc tca gat gga ctc ttc cag gcc atc      480
```

```
Thr Phe Pro Tyr Val Tyr Gly Phe Ser Asp Gly Leu Phe Gln Ala Ile
145                 150                 155                 160
ctg acc ttc cgc ctg acc ttc tgt aga tcc aat gtc atc aac cac ttc      528
Leu Thr Phe Arg Leu Thr Phe Cys Arg Ser Asn Val Ile Asn His Phe
                165                 170                 175
tac tgt gct gac ccg ccg ctc att aag ctt tct tgt tct gat act tat      576
Tyr Cys Ala Asp Pro Pro Leu Ile Lys Leu Ser Cys Ser Asp Thr Tyr
            180                 185                 190
gtc aaa gag cat gcc atg ttc ata tct gct ggc ttc aac ctc tcc agc      624
Val Lys Glu His Ala Met Phe Ile Ser Ala Gly Phe Asn Leu Ser Ser
        195                 200                 205
tcc ctc acc atc gtc ttg gtg tcc tat gcc ttc att ctt gct gcc atc      672
Ser Leu Thr Ile Val Leu Val Ser Tyr Ala Phe Ile Leu Ala Ala Ile
    210                 215                 220
ctc cgg atc aaa tca gca gag gga agg cac aag gca ttc tcc acc tgt      720
Leu Arg Ile Lys Ser Ala Glu Gly Arg His Lys Ala Phe Ser Thr Cys
225                 230                 235                 240
ggt tcc cat atg atg gct gtc acc ctg ttt tat ggg act ctc ttt tgc      768
Gly Ser His Met Met Ala Val Thr Leu Phe Tyr Gly Thr Leu Phe Cys
                245                 250                 255
atg tat ata aga cca cca aca gat aag act gtt gag gaa tct aaa ata      816
Met Tyr Ile Arg Pro Pro Thr Asp Lys Thr Val Glu Glu Ser Lys Ile
            260                 265                 270
ata gct gtc ttt tac acc ttt gtg agt ccg gta ctt aat cca ttg atc      864
Ile Ala Val Phe Tyr Thr Phe Val Ser Pro Val Leu Asn Pro Leu Ile
        275                 280                 285
tac agt ctg agg aat aaa gat gtg aag cag gcc ttg aag aat gtc ctg      912
Tyr Ser Leu Arg Asn Lys Asp Val Lys Gln Ala Leu Lys Asn Val Leu
    290                 295                 300
aga tga                                                             918
Arg  *
305
```

<210> 7
<211> 612
<212> DNA
<213> Homo sapiens

<400> 7

```
atg gtt aga gga aat tct act ttg gtg acg gaa ttt att ctc ttg gga       48
Met Val Arg Gly Asn Ser Thr Leu Val Thr Glu Phe Ile Leu Leu Gly
1               5                   10                  15
tta aag gat ctt cca gag ctt cag ccc atc ctc ttt gta ctg ttc ctg       96
Leu Lys Asp Leu Pro Glu Leu Gln Pro Ile Leu Phe Val Leu Phe Leu
            20                  25                  30
cta atc tac ctg atc act gtc ggg ggg aac ctt ggg atg ttg gtg ttg      144
Leu Ile Tyr Leu Ile Thr Val Gly Gly Asn Leu Gly Met Leu Val Leu
        35                  40                  45
atc agg ata gat tca cgc ctc cac acc ccc atg tat ttc ttt ctt gct      192
Ile Arg Ile Asp Ser Arg Leu His Thr Pro Met Tyr Phe Phe Leu Ala
    50                  55                  60
agt ttg tcc tgc ttg gat ttg tat tac tcc act aat gtg act ccc aag      240
Ser Leu Ser Cys Leu Asp Leu Tyr Tyr Ser Thr Asn Val Thr Pro Lys
65                  70                  75                  80
atg ttg gtg aac ttc ttc tca gac aag aaa gcc att tcc tat gct gct      288
Met Leu Val Asn Phe Phe Ser Asp Lys Lys Ala Ile Ser Tyr Ala Ala
                85                  90                  95
tgt tta gtc cag tgc tat ttt ttc att gct gtg gtg att act gaa tat      336
Cys Leu Val Gln Cys Tyr Phe Phe Ile Ala Val Val Ile Thr Glu Tyr
            100                 105                 110
tat atg cta gct gta atg gcc tat gat agg tat gtg gcc atc tgt aac      384
Tyr Met Leu Ala Val Met Ala Tyr Asp Arg Tyr Val Ala Ile Cys Asn
        115                 120                 125
```

```
cct ttg ctt tac agc agc aag atg tcc aaa ggg ctc tgt att cgc ctg      432
Pro Leu Leu Tyr Ser Ser Lys Met Ser Lys Gly Leu Cys Ile Arg Leu
    130                 135                 140
att gct ggt cca tat gtc tat ggg ttt ctt agt gga ctg atg gaa acc      480
Ile Ala Gly Pro Tyr Val Tyr Gly Phe Leu Ser Gly Leu Met Glu Thr
145                 150                 155                 160
atg tgg aca tac cac ttg acc ttc tgt ggc tcc aat atc att aat cac      528
Met Trp Thr Tyr His Leu Thr Phe Cys Gly Ser Asn Ile Ile Asn His
                165                 170                 175
ttc tac tgt gct gac cca ccc ctc atc cga ctt tcc tgc tct gac act      576
Phe Tyr Cys Ala Asp Pro Pro Leu Ile Arg Leu Ser Cys Ser Asp Thr
            180                 185                 190
ttc att aag gaa aca tcc atg ttt gtg gta gca tga                      612
Phe Ile Lys Glu Thr Ser Met Phe Val Val Ala  *
        195                 200

<210> 8
<211> 807
<212> DNA
<213> Homo sapiens

<400> 8
ttg ccc tca tcc agg cca acg ccc cgg ctc cac acg ccc atg tac ttt       48
Leu Pro Ser Ser Arg Pro Thr Pro Arg Leu His Thr Pro Met Tyr Phe
1               5                   10                  15
ttc ctg agc aac tta tcc ttt gtg gat ctg tgc ttc tct tcc aat gtg       96
Phe Leu Ser Asn Leu Ser Phe Val Asp Leu Cys Phe Ser Ser Asn Val
            20                  25                  30
act cca agg atg ctg gag att ttc ctt tca gag aag aaa agc att tcc      144
Thr Pro Arg Met Leu Glu Ile Phe Leu Ser Glu Lys Lys Ser Ile Ser
        35                  40                  45
tat cct gcc cgt ctt gtg cag tgt tac ctt ttt atc acc ttg gtc cac      192
Tyr Pro Ala Arg Leu Val Gln Cys Tyr Leu Phe Ile Thr Leu Val His
    50                  55                  60
gtt gag ctc tac atc ctg gct gtg atg gcc ttt gac cgg tac atg gcc      240
Val Glu Leu Tyr Ile Leu Ala Val Met Ala Phe Asp Arg Tyr Met Ala
65                  70                  75                  80
atc tgc aac cct ctg ctt tat ggc agc aga atg tcc aag agc gtg tgc      288
Ile Cys Asn Pro Leu Leu Tyr Gly Ser Arg Met Ser Lys Ser Val Cys
                85                  90                  95
tct ttc ctc atc aca gtg ctt tat gtg tat gga gca ctc act ggc ctg      336
Ser Phe Leu Ile Thr Val Leu Tyr Val Tyr Gly Ala Leu Thr Gly Leu
            100                 105                 110
atg gag act atg tgg acc tac aac cta gcc ttc tgt ggc ccc agt gaa      384
Met Glu Thr Met Trp Thr Tyr Asn Leu Ala Phe Cys Gly Pro Ser Glu
        115                 120                 125
att aat cac ttc tac tgt gtg gac cca cca ctg att aag ctg gct tgt      432
Ile Asn His Phe Tyr Cys Val Asp Pro Pro Leu Ile Lys Leu Ala Cys
    130                 135                 140
tct gac acc tac aac aag gag gtg tca atg ttt gtt gtg gct ggt ttc      480
Ser Asp Thr Tyr Asn Lys Glu Val Ser Met Phe Val Val Ala Gly Phe
145                 150                 155                 160
aac ttc act tat cct ctc ctt atc atc ctc att tcc tat ctc tac ata      528
Asn Phe Thr Tyr Pro Leu Leu Ile Ile Leu Ile Ser Tyr Leu Tyr Ile
                165                 170                 175
ttt cct gcc acc cta agg atc tgc tct aca gaa ggc agg cac aaa gct      576
Phe Pro Ala Thr Leu Arg Ile Cys Ser Thr Glu Gly Arg His Lys Ala
            180                 185                 190
ttt tct acc tgt ggc tcc cat ctg aca gcc gtt act att ttc tat tca      624
Phe Ser Thr Cys Gly Ser His Leu Thr Ala Val Thr Ile Phe Tyr Ser
        195                 200                 205
gct ctt ttc ttc atg tat ctc aga cgt cca tca gaa gag tcc atg gag      672
Ala Leu Phe Phe Met Tyr Leu Arg Arg Pro Ser Glu Glu Ser Met Glu
```

```
    210                 215                 220
cag ggg aaa atg gta gct gta ttt tat acc act gta atc ccc atg ttg      720
Gln Gly Lys Met Val Ala Val Phe Tyr Thr Thr Val Ile Pro Met Leu
225                 230                 235                 240
aat ccc atg atc tac agt ctg agg aac aaa gat gtg aaa gag gca tta      768
Asn Pro Met Ile Tyr Ser Leu Arg Asn Lys Asp Val Lys Glu Ala Leu
                245                 250                 255
tgc aaa gaa ctg ttc aaa aga aaa ttg ttt tct aaa taa                  807
Cys Lys Glu Leu Phe Lys Arg Lys Leu Phe Ser Lys  *
            260                 265
```

<210> 9
<211> 363
<212> DNA
<213> Homo sapiens

<400> 9

```
atg aga agg aac ttc acg ttg gtg act gag ttc att ctc ctg gga ctg       48
Met Arg Arg Asn Phe Thr Leu Val Thr Glu Phe Ile Leu Leu Gly Leu
1               5                   10                  15
acg aat cac cag gaa tta cag att ctc ctc ttc atg ctg ttt ctg gcc       96
Thr Asn His Gln Glu Leu Gln Ile Leu Leu Phe Met Leu Phe Leu Ala
            20                  25                  30
att tac atg gtc aca gtg gca ggg aat ctt agc atg att gcc ctc atc      144
Ile Tyr Met Val Thr Val Ala Gly Asn Leu Ser Met Ile Ala Leu Ile
        35                  40                  45
cag gcc aat gcc cgg ctc cac acg ccc atg tac ttt ttc ctg agc cac      192
Gln Ala Asn Ala Arg Leu His Thr Pro Met Tyr Phe Phe Leu Ser His
    50                  55                  60
tta tcc ttc ctg gat ctg tgc ttc tct tcc aat gtg acc cca aag atg      240
Leu Ser Phe Leu Asp Leu Cys Phe Ser Ser Asn Val Thr Pro Lys Met
65                  70                  75                  80
ctg gag att ttc ctt tca gag aag aaa agc att tcc tat cct gcc tgt      288
Leu Glu Ile Phe Leu Ser Glu Lys Lys Ser Ile Ser Tyr Pro Ala Cys
                85                  90                  95
ctt gtt cag tgt tac ctt tat atc atc ttg gta cac gtt gag atc tac      336
Leu Val Gln Cys Tyr Leu Tyr Ile Ile Leu Val His Val Glu Ile Tyr
            100                 105                 110
atc ctg gct gtg atg gcc ttt gac tag                                  363
Ile Leu Ala Val Met Ala Phe Asp  *
        115                 120
```

<210> 10
<211> 936
<212> DNA
<213> Homo sapiens

<400> 10

```
atg aga aga aac tgc acg ttg gtg act gag ttc att ctc ctg gga ctg       48
Met Arg Arg Asn Cys Thr Leu Val Thr Glu Phe Ile Leu Leu Gly Leu
1               5                   10                  15
acc agt cgc cgg gaa tta caa att ctc ctc ttc acg ctg ttt ctg gcc       96
Thr Ser Arg Arg Glu Leu Gln Ile Leu Leu Phe Thr Leu Phe Leu Ala
            20                  25                  30
att tac atg gtc acg gtg gca ggg aac ctt ggc atg att gtc ctc atc      144
Ile Tyr Met Val Thr Val Ala Gly Asn Leu Gly Met Ile Val Leu Ile
        35                  40                  45
cag gcc aac gcc tgg ctc cac atg ccc atg tac ttt ttc ctg agc cac      192
Gln Ala Asn Ala Trp Leu His Met Pro Met Tyr Phe Phe Leu Ser His
    50                  55                  60
tta tcc ttc gtg gat ctg tgc ttc tct tcc aat gtg act cca aag atg      240
Leu Ser Phe Val Asp Leu Cys Phe Ser Ser Asn Val Thr Pro Lys Met
65                  70                  75                  80
```

```
ctg gag att ttc ctt tca gag aag aaa agc att tcc tat cct gcc tgt      288
Leu Glu Ile Phe Leu Ser Glu Lys Lys Ser Ile Ser Tyr Pro Ala Cys
                85                  90                  95
ctt gtg cag tgt tac ctt ttt atc gcc ttg gtc cat gtt gag atc tac      336
Leu Val Gln Cys Tyr Leu Phe Ile Ala Leu Val His Val Glu Ile Tyr
            100                 105                 110
atc ctg gct gtg atg gcc ttt gac cgg tac atg gcc atc tgc aac cct      384
Ile Leu Ala Val Met Ala Phe Asp Arg Tyr Met Ala Ile Cys Asn Pro
        115                 120                 125
ctg ctt tat ggc agc aga atg tcc aag agt gtg tgc tcc ttc ctc atc      432
Leu Leu Tyr Gly Ser Arg Met Ser Lys Ser Val Cys Ser Phe Leu Ile
    130                 135                 140
acg gtg cct tat gtg tat gga gcg ctc act ggc ctg atg gag acc atg      480
Thr Val Pro Tyr Val Tyr Gly Ala Leu Thr Gly Leu Met Glu Thr Met
145                 150                 155                 160
tgg acc tac aac cta gcc ttc tgt ggc ccc aat gaa att aat cac ttc      528
Trp Thr Tyr Asn Leu Ala Phe Cys Gly Pro Asn Glu Ile Asn His Phe
                165                 170                 175
tac tgt gcg gac cca cca ctg att aag ctg gct tgt tct gac acc tac      576
Tyr Cys Ala Asp Pro Pro Leu Ile Lys Leu Ala Cys Ser Asp Thr Tyr
            180                 185                 190
aac aag gag ttg tca atg ttt att gtg gct ggc tgg aac ctt tct ttt      624
Asn Lys Glu Leu Ser Met Phe Ile Val Ala Gly Trp Asn Leu Ser Phe
        195                 200                 205
tct ctc ttc atc ata tgt att tcc tac ctt tac att ttc cct gct att      672
Ser Leu Phe Ile Ile Cys Ile Ser Tyr Leu Tyr Ile Phe Pro Ala Ile
    210                 215                 220
tta aag att cgc tct aca gag ggc agg caa aaa gct ttt tct acc tgt      720
Leu Lys Ile Arg Ser Thr Glu Gly Arg Gln Lys Ala Phe Ser Thr Cys
225                 230                 235                 240
ggc tcc cat ctg aca gct gtc act ata ttc tat gca acc ctt ttc ttc      768
Gly Ser His Leu Thr Ala Val Thr Ile Phe Tyr Ala Thr Leu Phe Phe
                245                 250                 255
atg tat ctc aga ccc ccc tca aag gaa tct gtt gaa cag ggt aaa atg      816
Met Tyr Leu Arg Pro Pro Ser Lys Glu Ser Val Glu Gln Gly Lys Met
            260                 265                 270
gta gct gta ttt tat acc aca gta atc cct atg ctg aac ctt ata att      864
Val Ala Val Phe Tyr Thr Thr Val Ile Pro Met Leu Asn Leu Ile Ile
        275                 280                 285
tat agc ctt aga aat aaa aat gta aaa gaa gca tta atc aaa gag ctg      912
Tyr Ser Leu Arg Asn Lys Asn Val Lys Glu Ala Leu Ile Lys Glu Leu
    290                 295                 300
tca atg aag ata tac ttt tct taa                                     936
Ser Met Lys Ile Tyr Phe Ser  *
305                 310
```

<210> 11
<211> 180
<212> DNA
<213> Homo sapiens

<400> 11

```
atg tcc aga aga aac tat act gaa ctg aca gaa ttt gtt ctc ttg ggt       48
Met Ser Arg Arg Asn Tyr Thr Glu Leu Thr Glu Phe Val Leu Leu Gly
1               5                   10                  15
cta aca agc cgt cca gag ctg cga gtt gct ttc ttg gca ctg ttc ctt       96
Leu Thr Ser Arg Pro Glu Leu Arg Val Ala Phe Leu Ala Leu Phe Leu
            20                  25                  30
ttt gtc tac ata gcc act gtg gta gga aac ttg ggg atg att att tta      144
Phe Val Tyr Ile Ala Thr Val Val Gly Asn Leu Gly Met Ile Ile Leu
        35                  40                  45
atc aaa gtt gat tct cga ctt cac act ccc atg taa                      180
Ile Lys Val Asp Ser Arg Leu His Thr Pro Met  *
```

50 55 60

<210> 12
<211> 64
<212> PRT
<213> Homo sapiens

<400> 12
Met Ser Phe Leu Ile Arg Ser Asp Ser Thr Leu His Thr Pro Met Cys
1 5 10 15
Leu Phe Leu Ser His Leu Ser Phe Val Asp Leu Tyr Tyr Ala Thr Asn
20 25 30
Ala Thr Pro Pro Met Leu Val Asn Phe Phe Phe Pro Arg Glu Lys Pro
35 40 45
Phe Pro Leu Leu Val Ala Leu Ser Asn Phe Thr Phe Ser Leu His Trp
50 55 60

<210> 13
<211> 315
<212> PRT
<213> Homo sapiens

<400> 13
Met Phe Ser Pro Asn His Thr Ile Val Thr Glu Phe Ile Leu Leu Gly
1 5 10 15
Leu Thr Asp Asp Pro Val Leu Glu Lys Ile Leu Phe Gly Val Phe Leu
20 25 30
Ala Ile Tyr Leu Ile Thr Leu Ala Gly Asn Leu Cys Met Ile Leu Leu
35 40 45
Ile Arg Thr Asn Ser His Leu Gln Thr Pro Met Tyr Phe Phe Leu Gly
50 55 60
His Leu Ser Phe Val Asp Ile Cys Tyr Ser Ser Asn Val Thr Pro Asn
65 70 75 80
Met Leu His Asn Phe Leu Ser Glu Gln Lys Thr Ile Ser Tyr Ala Gly
85 90 95
Cys Phe Thr Gln Cys Leu Leu Phe Ile Ala Leu Val Ile Thr Glu Phe
100 105 110
Tyr Phe Leu Ala Ser Met Ala Leu Asp Arg Tyr Val Ala Ile Cys Ser
115 120 125
Pro Leu His Tyr Ser Ser Arg Met Ser Lys Asn Ile Cys Ile Ser Leu
130 135 140
Val Thr Val Pro Tyr Met Tyr Gly Phe Leu Asn Gly Leu Ser Gln Thr
145 150 155 160
Leu Leu Thr Phe His Leu Ser Phe Cys Gly Ser Leu Glu Ile Asn His
165 170 175
Phe Tyr Cys Ala Asp Pro Pro Leu Ile Met Leu Ala Cys Ser Asp Thr
180 185 190
Arg Val Lys Lys Met Ala Met Phe Val Val Ala Gly Phe Thr Leu Ser
195 200 205
Ser Ser Leu Phe Ile Ile Leu Leu Ser Tyr Leu Phe Ile Phe Ala Ala
210 215 220
Ile Phe Arg Ile Arg Ser Ala Glu Gly Arg His Lys Ala Phe Ser Thr
225 230 235 240
Cys Ala Ser His Leu Thr Ile Val Thr Leu Phe Tyr Gly Thr Leu Phe
245 250 255
Cys Met Tyr Val Arg Pro Pro Ser Glu Lys Ser Val Glu Glu Ser Lys
260 265 270
Ile Thr Ala Val Phe Tyr Thr Phe Leu Thr Pro Met Leu Asn Pro Leu
275 280 285
Ile Tyr Ser Leu Arg Asn Thr Asp Val Ile Leu Ala Met Gln Gln Met
290 295 300
Ile Arg Gly Lys Ser Phe His Lys Ile Ala Val
305 310 315

<210> 14
<211> 172
<212> PRT
<213> Homo sapiens

<400> 14
Met Leu Lys Lys Asn His Thr Ala Val Thr Glu Phe Val Leu Leu Gly
1 5 10 15
Leu Thr Asp Arg Ala Glu Leu Gln Ser Leu Leu Phe Val Val Phe Leu
20 25 30
Val Ile Tyr Leu Ile Thr Val Ile Gly Asn Val Ser Met Ile Leu Leu
35 40 45
Ile Arg Ser Asp Ser Thr Leu His Thr Pro Met Tyr Phe Phe Leu Ser
50 55 60
His Leu Ser Phe Val Asp Leu Cys Tyr Thr Thr Asn Val Thr Pro Gln
65 70 75 80
Met Leu Val Asn Phe Leu Ser Lys Arg Lys Thr Ile Ser Phe Ile Gly
85 90 95
Cys Phe Ile Gln Phe His Phe Phe Ile Ala Leu Val Ile Thr Asp Tyr
100 105 110
Tyr Met Leu Thr Val Met Ala Tyr Asp Arg Tyr Met Ala Ile Cys Lys
115 120 125
Pro Leu Leu Tyr Gly Ser Lys Met Thr Arg Cys Val Cys Leu Cys Leu
130 135 140
Ala Ala Ala Pro Tyr Ile Tyr Gly Phe Ala Asn Gly Leu Ser Thr Asp
145 150 155 160
His Pro Asp Ala Ser Ser Val Leu Leu Trp Thr Gln
165 170

<210> 15
<211> 315
<212> PRT
<213> Homo sapiens

<400> 15
Met Leu Ser Pro Asn His Thr Ile Val Thr Glu Phe Ile Leu Leu Gly
1 5 10 15
Leu Thr Asp Asp Pro Val Leu Glu Lys Ile Leu Phe Gly Val Phe Leu
20 25 30
Ala Ile Tyr Leu Ile Thr Leu Ala Gly Asn Leu Cys Met Ile Leu Leu
35 40 45
Ile Arg Thr Asn Ser Gln Leu Gln Thr Pro Met Tyr Phe Phe Leu Gly
50 55 60
His Leu Ser Phe Leu Asp Ile Cys Tyr Ser Ser Asn Val Thr Pro Asn
65 70 75 80
Met Leu His Asn Phe Leu Ser Glu Gln Lys Thr Ile Ser Tyr Ala Gly
85 90 95
Cys Phe Thr Gln Cys Leu Leu Phe Ile Ala Leu Val Ile Thr Glu Phe
100 105 110
Tyr Phe Leu Ala Ser Met Ala Leu Asp Arg Tyr Val Ala Ile Cys Ser
115 120 125
Pro Leu His Tyr Ser Ser Arg Met Ser Lys Asn Ile Cys Ile Ser Leu
130 135 140
Val Thr Val Pro Tyr Met Tyr Gly Phe Leu Asn Gly Leu Ser Gln Thr
145 150 155 160
Leu Leu Thr Phe His Leu Ser Phe Cys Gly Ser Leu Glu Ile Asn His
165 170 175
Phe Tyr Cys Ala Asp Pro Pro Leu Ile Met Leu Ala Cys Ser Asp Thr
180 185 190
Arg Val Lys Lys Met Ala Met Phe Val Val Ala Gly Phe Thr Leu Ser
195 200 205
Ser Ser Leu Phe Ile Ile Leu Leu Ser Tyr Leu Phe Ile Phe Ala Ala

```
    210                 215                 220
Ile Phe Arg Ile Arg Ser Ala Glu Gly Arg His Lys Ala Phe Ser Thr
225                 230                 235                 240
Cys Ala Ser His Leu Thr Ile Val Thr Leu Phe Tyr Gly Thr Leu Phe
                245                 250                 255
Cys Met Tyr Val Arg Pro Pro Ser Glu Lys Ser Val Glu Glu Ser Lys
            260                 265                 270
Ile Ile Ala Val Phe Tyr Thr Phe Leu Ser Pro Met Leu Asn Pro Leu
        275                 280                 285
Ile Tyr Ser Leu Arg Asn Arg Asp Val Ile Leu Ala Ile Gln Gln Met
    290                 295                 300
Ile Arg Gly Lys Ser Phe Cys Lys Ile Ala Val
305                 310                 315

<210> 16
<211> 305
<212> PRT
<213> Homo sapiens

<400> 16
Met Ser Asn Thr Asn Gly Ser Ala Ile Thr Glu Phe Ile Leu Leu Gly
1               5                   10                  15
Leu Thr Asp Cys Pro Glu Leu Gln Ser Leu Leu Phe Val Leu Phe Leu
            20                  25                  30
Val Val Tyr Leu Val Thr Leu Leu Gly Asn Leu Gly Met Ile Met Leu
        35                  40                  45
Met Arg Leu Asp Ser Arg Leu His Thr Pro Met Tyr Phe Phe Leu Thr
    50                  55                  60
Asn Leu Ala Phe Val Asp Leu Cys Tyr Thr Ser Asn Ala Thr Pro Gln
65                  70                  75                  80
Met Ser Thr Asn Ile Val Ser Glu Lys Thr Ile Ser Phe Ala Gly Cys
                85                  90                  95
Phe Thr Gln Cys Tyr Ile Phe Ile Ala Leu Leu Leu Thr Glu Phe Tyr
            100                 105                 110
Met Leu Ala Ala Met Ala Tyr Asp Arg Tyr Val Ala Ile Tyr Asp Pro
        115                 120                 125
Leu Arg Tyr Ser Val Lys Thr Ser Arg Arg Val Cys Ile Cys Leu Ala
    130                 135                 140
Thr Phe Pro Tyr Val Tyr Gly Phe Ser Asp Gly Leu Phe Gln Ala Ile
145                 150                 155                 160
Leu Thr Phe Arg Leu Thr Phe Cys Arg Ser Asn Val Ile Asn His Phe
                165                 170                 175
Tyr Cys Ala Asp Pro Pro Leu Ile Lys Leu Ser Cys Ser Asp Thr Tyr
            180                 185                 190
Val Lys Glu His Ala Met Phe Ile Ser Ala Gly Phe Asn Leu Ser Ser
        195                 200                 205
Ser Leu Thr Ile Val Leu Val Ser Tyr Ala Phe Ile Leu Ala Ala Ile
    210                 215                 220
Leu Arg Ile Lys Ser Ala Glu Gly Arg His Lys Ala Phe Ser Thr Cys
225                 230                 235                 240
Gly Ser His Met Met Ala Val Thr Leu Phe Tyr Gly Thr Leu Phe Cys
                245                 250                 255
Met Tyr Ile Arg Pro Pro Thr Asp Lys Thr Val Glu Glu Ser Lys Ile
            260                 265                 270
Ile Ala Val Phe Tyr Thr Phe Val Ser Pro Val Leu Asn Pro Leu Ile
        275                 280                 285
Tyr Ser Leu Arg Asn Lys Asp Val Lys Gln Ala Leu Lys Asn Val Leu
    290                 295                 300
Arg
305

<210> 17
<211> 203
```

<212> PRT
<213> Homo sapiens

<400> 17
Met Val Arg Gly Asn Ser Thr Leu Val Thr Glu Phe Ile Leu Leu Gly
1 5 10 15
Leu Lys Asp Leu Pro Glu Leu Gln Pro Ile Leu Phe Val Leu Phe Leu
20 25 30
Leu Ile Tyr Leu Ile Thr Val Gly Gly Asn Leu Gly Met Leu Val Leu
35 40 45
Ile Arg Ile Asp Ser Arg Leu His Thr Pro Met Tyr Phe Phe Leu Ala
50 55 60
Ser Leu Ser Cys Leu Asp Leu Tyr Tyr Ser Thr Asn Val Thr Pro Lys
65 70 75 80
Met Leu Val Asn Phe Phe Ser Asp Lys Lys Ala Ile Ser Tyr Ala Ala
85 90 95
Cys Leu Val Gln Cys Tyr Phe Phe Ile Ala Val Val Ile Thr Glu Tyr
100 105 110
Tyr Met Leu Ala Val Met Ala Tyr Asp Arg Tyr Val Ala Ile Cys Asn
115 120 125
Pro Leu Leu Tyr Ser Ser Lys Met Ser Lys Gly Leu Cys Ile Arg Leu
130 135 140
Ile Ala Gly Pro Tyr Val Tyr Gly Phe Leu Ser Gly Leu Met Glu Thr
145 150 155 160
Met Trp Thr Tyr His Leu Thr Phe Cys Gly Ser Asn Ile Ile Asn His
165 170 175
Phe Tyr Cys Ala Asp Pro Pro Leu Ile Arg Leu Ser Cys Ser Asp Thr
180 185 190
Phe Ile Lys Glu Thr Ser Met Phe Val Val Ala
195 200

<210> 18
<211> 268
<212> PRT
<213> Homo sapiens

<400> 18
Leu Pro Ser Ser Arg Pro Thr Pro Arg Leu His Thr Pro Met Tyr Phe
1 5 10 15
Phe Leu Ser Asn Leu Ser Phe Val Asp Leu Cys Phe Ser Ser Asn Val
20 25 30
Thr Pro Arg Met Leu Glu Ile Phe Leu Ser Glu Lys Lys Ser Ile Ser
35 40 45
Tyr Pro Ala Arg Leu Val Gln Cys Tyr Leu Phe Ile Thr Leu Val His
50 55 60
Val Glu Leu Tyr Ile Leu Ala Val Met Ala Phe Asp Arg Tyr Met Ala
65 70 75 80
Ile Cys Asn Pro Leu Leu Tyr Gly Ser Arg Met Ser Lys Ser Val Cys
85 90 95
Ser Phe Leu Ile Thr Val Leu Tyr Val Tyr Gly Ala Leu Thr Gly Leu
100 105 110
Met Glu Thr Met Trp Thr Tyr Asn Leu Ala Phe Cys Gly Pro Ser Glu
115 120 125
Ile Asn His Phe Tyr Cys Val Asp Pro Pro Leu Ile Lys Leu Ala Cys
130 135 140
Ser Asp Thr Tyr Asn Lys Glu Val Ser Met Phe Val Val Ala Gly Phe
145 150 155 160
Asn Phe Thr Tyr Pro Leu Leu Ile Ile Leu Ile Ser Tyr Leu Tyr Ile
165 170 175
Phe Pro Ala Thr Leu Arg Ile Cys Ser Thr Glu Gly Arg His Lys Ala
180 185 190
Phe Ser Thr Cys Gly Ser His Leu Thr Ala Val Thr Ile Phe Tyr Ser
195 200 205

```
Ala Leu Phe Phe Met Tyr Leu Arg Arg Pro Ser Glu Glu Ser Met Glu
    210                 215                 220
Gln Gly Lys Met Val Ala Val Phe Tyr Thr Thr Val Ile Pro Met Leu
225                 230                 235                 240
Asn Pro Met Ile Tyr Ser Leu Arg Asn Lys Asp Val Lys Glu Ala Leu
                245                 250                 255
Cys Lys Glu Leu Phe Lys Arg Lys Leu Phe Ser Lys
            260                 265
```

<210> 19
<211> 120
<212> PRT
<213> Homo sapiens

<400> 19

```
Met Arg Arg Asn Phe Thr Leu Val Thr Glu Phe Ile Leu Leu Gly Leu
1               5                   10                  15
Thr Asn His Gln Glu Leu Gln Ile Leu Leu Phe Met Leu Phe Leu Ala
            20                  25                  30
Ile Tyr Met Val Thr Val Ala Gly Asn Leu Ser Met Ile Ala Leu Ile
        35                  40                  45
Gln Ala Asn Ala Arg Leu His Thr Pro Met Tyr Phe Phe Leu Ser His
    50                  55                  60
Leu Ser Phe Leu Asp Leu Cys Phe Ser Ser Asn Val Thr Pro Lys Met
65                  70                  75                  80
Leu Glu Ile Phe Leu Ser Glu Lys Lys Ser Ile Ser Tyr Pro Ala Cys
                85                  90                  95
Leu Val Gln Cys Tyr Leu Tyr Ile Ile Leu Val His Val Glu Ile Tyr
            100                 105                 110
Ile Leu Ala Val Met Ala Phe Asp
        115                 120
```

<210> 20
<211> 311
<212> PRT
<213> Homo sapiens

<400> 20

```
Met Arg Arg Asn Cys Thr Leu Val Thr Glu Phe Ile Leu Leu Gly Leu
1               5                   10                  15
Thr Ser Arg Arg Glu Leu Gln Ile Leu Leu Phe Thr Leu Phe Leu Ala
            20                  25                  30
Ile Tyr Met Val Thr Val Ala Gly Asn Leu Gly Met Ile Val Leu Ile
        35                  40                  45
Gln Ala Asn Ala Trp Leu His Met Pro Met Tyr Phe Phe Leu Ser His
    50                  55                  60
Leu Ser Phe Val Asp Leu Cys Phe Ser Ser Asn Val Thr Pro Lys Met
65                  70                  75                  80
Leu Glu Ile Phe Leu Ser Glu Lys Lys Ser Ile Ser Tyr Pro Ala Cys
                85                  90                  95
Leu Val Gln Cys Tyr Leu Phe Ile Ala Leu Val His Val Glu Ile Tyr
            100                 105                 110
Ile Leu Ala Val Met Ala Phe Asp Arg Tyr Met Ala Ile Cys Asn Pro
        115                 120                 125
Leu Leu Tyr Gly Ser Arg Met Ser Lys Ser Val Cys Ser Phe Leu Ile
    130                 135                 140
Thr Val Pro Tyr Val Tyr Gly Ala Leu Thr Gly Leu Met Glu Thr Met
145                 150                 155                 160
Trp Thr Tyr Asn Leu Ala Phe Cys Gly Pro Asn Glu Ile Asn His Phe
                165                 170                 175
Tyr Cys Ala Asp Pro Pro Leu Ile Lys Leu Ala Cys Ser Asp Thr Tyr
            180                 185                 190
Asn Lys Glu Leu Ser Met Phe Ile Val Ala Gly Trp Asn Leu Ser Phe
```

```
        195                 200                 205
Ser Leu Phe Ile Ile Cys Ile Ser Tyr Leu Tyr Ile Phe Pro Ala Ile
    210                 215                 220
Leu Lys Ile Arg Ser Thr Glu Gly Arg Gln Lys Ala Phe Ser Thr Cys
225                 230                 235                 240
Gly Ser His Leu Thr Ala Val Thr Ile Phe Tyr Ala Thr Leu Phe Phe
                245                 250                 255
Met Tyr Leu Arg Pro Pro Ser Lys Glu Ser Val Glu Gln Gly Lys Met
            260                 265                 270
Val Ala Val Phe Tyr Thr Thr Val Ile Pro Met Leu Asn Leu Ile Ile
        275                 280                 285
Tyr Ser Leu Arg Asn Lys Asn Val Lys Glu Ala Leu Ile Lys Glu Leu
    290                 295                 300
Ser Met Lys Ile Tyr Phe Ser
305                 310
```

<210> 21
<211> 59
<212> PRT
<213> Homo sapiens

<400> 21

```
Met Ser Arg Arg Asn Tyr Thr Glu Leu Thr Glu Phe Val Leu Leu Gly
1               5                   10                  15
Leu Thr Ser Arg Pro Glu Leu Arg Val Ala Phe Leu Ala Leu Phe Leu
            20                  25                  30
Phe Val Tyr Ile Ala Thr Val Val Gly Asn Leu Gly Met Ile Ile Leu
        35                  40                  45
Ile Lys Val Asp Ser Arg Leu His Thr Pro Met
    50                  55
```

<210> 22
<211> 30
<212> DNA
<213> Artificial Sequence

<400> 22
cctggagggt ttcaaaggct gatactttag 30

<210> 23
<211> 26
<212> DNA
<213> Artificial Sequence

<400> 23
ctccagcctg agcaacagag caatac 26

<210> 24
<211> 30
<212> DNA
<213> Artificial Sequence

<400> 24
ctcacattca ttgttcttca cagacccagc 30

<210> 25
<211> 24
<212> DNA
<213> Artificial Sequence

<400> 25
ccctgctggg atctggatca agac 24

<210> 26
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> sequencing oligonucleotide PrimerPU

<400> 26
tgtaaaacga cggccagt 18

<210> 27
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> sequencing oligonucleotide PrimerRP

<400> 27
caggaaacag ctatgacc 18